

Pursuing diversified universities

The Australian government's proposed university policy increases competition and market forces in the sector, while raising justified concerns about independence and infrastructure. But the goal of diversifying higher education is appropriate.

Establishing suitable levels of access to higher education and funding academic research infrastructure are challenges that face many countries. Australia has long been in the vanguard of providing publicly funded student loans, while funds for its laboratories are all too scarce. So a significant change in its approach to higher education is worth watching. The Australian federal government has now unveiled its plan for the most significant reform of the nation's universities in more than a decade. The blueprint seems to reflect the government's philosophy of 'the user pays' and the promotion of elite institutions. The good news is that the universities will probably receive more money. But there are concerns about growing student debt and potential government influence over university curricula.

The blueprint allows universities to ratchet up student fees and increase the proportion of full-fee-paying students. Currently, universities enroll students through two routes. For government-subsidized places, students contribute a proportion of the costs, payment of which can be delayed until they have graduated and reach a threshold salary in the workplace. Alternatively, university places can be filled with full-fee-paying students, who, in some instances, can gain entry on lower academic scores than those competing for subsidized places. Under the new scheme, the proportion of university places available for full-fee-paying students will rise from 25% to 50%. The fee contributions demanded of government-subsidized students can be increased to 30% above current levels.

It is not surprising that many university chiefs have embraced this partial fee deregulation, as universities are parched from years of declining government funds. Alarming, the modest additional government contribution to courses — A\$404 million (US\$267 million) over four years from 2005 — has strings attached. The universities must comply with reforms including individual workplace agree-

ments, restrictions on industrial action and the abolition of compulsory student unions. Furthermore, the government will negotiate contributions to each individual institution, and it is unclear whether it will play a controlling or a consultative role in determining the type of courses and number of places offered by different universities.

The reform is a realization of the government's ambition to propel a few Australian universities into the upper echelons of global tertiary institutions, attracting elite academics and conducting world-class teaching and research. Beneath these will be a spectrum of less prestigious, more vocationally oriented universities, which are likely to be more focused on teaching than research. Some universities are likely to focus on their strengths in order to attract fee-paying students, leading to specialization in some parts of the sector. This is desirable: in a country with a limited population and resources, the idea of all universities trying to be all things to all people is hopelessly idealistic.

An egregious omission from the reform package is any mention of further infrastructure support for universities, many of which have to use teaching money to support their research, as Australian grants rarely cover overheads. The government must redress this when it completes its ongoing audit of the nation's scientific landscape.

But is further exposing the universities to market forces the best way of obtaining excellence? For a nation hoping to benefit economically and socially from the fruits of world-class university research and teaching, more substantial government investment would have been appropriate. Driving universities to compete for fee-paying students runs the risk of reshaping universities as sites of vocational training rather than as places of higher learning. Without striking the right balance between public funding and student contributions, the Australian government may severely blunt the leading edge, both in education and research, that its best universities represent. ■

Recovering from cultural devastation

Leaders of the world's scientific community must act with more speed and determination to help reconstruct Iraq.

In central Baghdad, Iraq's National Library and National Archives lie ransacked and burnt to the ground, under the noses of coalition troops. The Bush administration's disrespect for a basic precept of the prosecution of war — the preservation of cultural heritage — has led to the unnecessary loss of much of the record of 10,000 years of history. Baghdad's Iraqi National Museum, the world's finest collection on earliest human civilization, has been plundered. So too have its universities (see page 468).

Military schools will no doubt teach that the United States' technological superiority in battle significantly reduced 'collateral damage' in terms of civilian lives. They might also discuss the 1954 Hague Convention for the Protection of Cultural Property in the Event of Armed Conflict, which the United States has not ratified.

With lives being lost, why should one care about science and culture? Because wars end, and shattered lives, communities and societies must be rebuilt. Amid the pressing humanitarian needs, the international community must help to reconstruct the training of the next

generation to lead Iraq. The immediate priority is to put the universities back on their feet. International scientific collaboration must then use the full 'shock and awe' of civilian science and technology to protect public health, secure water resources, and rebuild infrastructure and the economy. So far, the international community has been mute in this regard, and the United Nations agency charged with science, UNESCO, has been slow and ineffective.

That must change, and fast. The top priority is to establish the state of infrastructure and the most pressing needs. A conference, perhaps under the auspices of UNESCO, could bring together Iraqi scientists and outside organizations. At the same time, individual scientists and their institutions can reach out to their counterparts in Iraq.

The Bush leadership intends to purge science of Ba'ath Party members, but it should remember that despite widespread nepotism at senior levels, allegiance was often token. Many decent scientists who can help to rebuild Iraq were Ba'ath Party members out of pragmatism. They must be recruited for the reconstruction that lies ahead. ■





Out of the ashes
Iraq's universities
begin to pick up
the pieces
p468



Nuclear option
Beefed-up bunker
busters fail to win
favour with experts
p469



In a spin
Biologists hatch
ambitious plan
to track viruses
p471



Tall story
Giant flower
breaks the
blooming record
p472

Virus detectives seek source of SARS in China's wild animals

David Cyranoski, Tokyo,
and Alison Abbott, Munich

Researchers investigating the source of severe acute respiratory syndrome (SARS) have turned their attention to the wild-animal markets of southern China. The move follows reports that workers and animals at the markets show high rates of infection with coronaviruses, the family to which the virus believed to cause SARS belongs.

The possible link to wild animals emerged on 23 May, when a team from the University of Hong Kong revealed that a coronavirus resembling the SARS virus had been isolated from six masked palm civets (*Paguma larvata*) and a raccoon dog (*Nyctereutes procyonoides*) in a market in Shenzhen, in the southern Chinese province of Guangdong. Antibodies against the virus were also found in a Chinese ferret badger (*Melogale moschata*) from the same market.

Although the virus is not the same as that believed to cause SARS — a member of the Hong Kong team describes it as “genetically very close, but not identical” — five out of the ten civet handlers at the market had antibodies against the SARS virus in their blood.



In the frame: a virus similar to that thought to cause SARS has been found in some masked palm civets.

A Chinese government team has since released results showing that 66 out of 508 animal handlers tested at markets in Guangdong had antibodies against the SARS virus.

Microbiologist Kwok-yung Yuen, who led the Hong Kong team, notes that in the

general population, the level of antibodies against the SARS virus is much lower. “This suggests that the virus is jumping from wild animals to humans,” he says.

Because the animal virus is similar to the SARS virus, but different from other coronaviruses, it is now a prime suspect in the hunt for the origins of SARS. The likelihood that the virus is moving the other way — from humans to animals — is diminished by ongoing work on genetic sequences of the two viruses. These analyses suggest that the animal version has an extra stretch of 29 nucleotide bases. “Viruses tend not to gain stretches of nucleotides when they jump across species,” says Klaus Stöhr, a SARS expert at the World Health Organization (WHO) in Geneva.

But uncertainties remain over the exact source of the virus. Only six civets were present in the market, and the fact that they all had the virus suggests that they were infected recently. “They could have got it from another animal during transport to the markets,” says Zhang-liang Chen, a molecular biologist and president of the China Agricultural University in Beijing. Chen’s group did not find the SARS virus in samples from eight civets taken from other markets and in the wild in Guangdong.

BSE case rattles Canadian officials

Hannah Hoag

The life and times of a cow from the Canadian province of Alberta are currently the subject of intense scrutiny, as officials struggle to understand how it became infected with bovine spongiform encephalopathy (BSE).

Canadian agriculture minister Lyle Vancil announced on 20 May that the animal, which was slaughtered in January, had tested positive for BSE. Within hours, the US Department of Agriculture (USDA) banned Canadian beef imports. As of 26 May, 17 herds in Alberta, Saskatchewan and British Columbia had been quarantined.

How the cow contracted the disease continues to perplex officials. Experts say that it is likely that the animal was infected through feed made from the remains of

cattle with the disease. Canada banned the use of animal remains in cattle feed in 1997, around the time that the cow was born.

Officials now have to determine what proportion of the country’s cattle herd, if any, has the disease. Canada continues to allow cattle remains to be used in feed for animals such as chickens, raising the possibility that infected feed could be finding its way to cattle.

But experts are betting against a large-scale outbreak or any danger to consumers. About one million cows are imported into the United States from Canada every year. Lisa Ferguson, who works on BSE-like diseases at the USDA, says that the United States runs BSE tests on these cattle, and would have detected any sign of an epidemic.

Iraqi science faces lonely road to recovery

Declan Butler

Iraq's beleaguered universities took their first tentative steps towards revival last week by electing interim leaders. But so far the institutions have received precious few offers of outside help in repairing the destruction they suffered at the hands of looters.

Rebuilding Iraq's science, technology and academic base, already weakened by years of economic sanctions, will be a huge task. Looters ransacked and burned parts of Baghdad and Basra universities, stripping them of everything from computers and pipettes to chairs and doors. Similar events occurred at other Baghdad institutions, including Iraq's National Library, which housed more than 500,000 books, and the Al-Awqaf library, home to some 6,000 Islamic manuscripts.

With sanctions against Iraq now lifted, and oil revenue available to pay for reconstruction, these institutions could, in theory, be rebuilt. But officials such as Sami al-Mudhafar, the biochemist elected last week as acting president of Baghdad University, do not seem to have received any substantial offer of help.

An official at the US Department of Defense's Office for Reconstruction and Humanitarian Assistance, which is overseeing the rebuilding of much of Iraq's infrastructure, says that the office's educational focus will be on "primary and secondary schools, rather than universities". The United Nations Educational, Scientific and Cultural



Building for the future? US troops guarded Iraqi universities during their leadership elections.

Organization (UNESCO) last week pledged to help to rebuild Iraq's cultural and educational institutions, but made no specific mention of science. And officials at the European Commission gave no details of any plan to help Iraq's scientific bodies.

Smaller bodies are currently focusing on assessing the extent of the damage before deciding what action to take. The Blue

Shield, a counterpart of the Red Cross that aims to protect libraries, archives and cultural heritage in times of war and natural disaster, is posting information on damage on its website.

And Henry Wright, an archaeologist at the University of Michigan in Ann Arbor, is due to return this week from a fact-finding mission in Iraq for the US National Academy of Sciences. "We are trying to figure out what we can do to help rebuild their universities and scientific communities," says Bruce Alberts, the academy's president.

If reconstruction funds are channelled towards universities, other groups will be able to provide outside support. The Health InterNetwork, a joint venture between the United Nations and publishers that is aimed at making electronic medical journals available to scientists in developing nations for free or at reduced prices, is discussing the rapid inclusion of Iraq in the scheme.

The US National Science Foundation operates an established scheme for funding research in the Middle East. Osman Shinaishin, the programme's manager, says that he began to receive proposals from US researchers for scientific and engineering projects with Iraqi scientists even before the war ended. "If a proposal reaches me it will go through a technical-merit assessment, and will then be funded as long as travel is safe," Shinaishin says. ■

ALI HAIDER/AP

Anthropologists cast doubt on human DNA evidence

Alison Abbott, Munich

When Giorgio Bertorelle published his latest findings on ancient mitochondrial DNA earlier this month, he claimed that his work added yet more weight to the idea that modern humans evolved independently of Neanderthals. But his paper may have a different significance — experts say that the techniques involved, which are also used by other groups, are dogged by problems so fundamental that they can never be used to draw conclusions about the evolution of modern humans.

Bertorelle, based at the University of Ferrara in Italy, claimed to have extracted mitochondrial DNA, which evolves independently of chromosomal DNA, from the bones of Cro-Magnons, *Homo sapiens* from around 25,000 years ago. The extracted sequences are strikingly similar to modern human DNA, and are very different from those of Neanderthal specimens (D. Caramelli *et al. Proc. Natl Acad. Sci. USA* 100, 6593–6597; 2003).

But experts question whether Bertorelle,

and others who are attempting similar experiments, can rule out contamination with material from modern humans. "Cro-Magnon DNA is so similar to modern human DNA that there is no way to say whether what has been seen is real," says Svante Pääbo, a director at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

Alan Cooper, an evolutionary anthropologist at the University of Oxford, UK, published the accepted standards of contamination avoidance in 2000 (A. Cooper and H. N. Poinar *Science* 289, 1139; 2000). He has deliberately contaminated teeth and bones with viral DNA by soaking them in liquid containing the virus. Cooper says that not even extreme decontamination methods, such as application of acid or bleach, can eradicate such DNA completely.

Bertorelle agrees that it is impossible to rule out contamination, but says he used all of the standard precautions in his study. "We repeated the procedure on a bone from a horse taken from the same layer of the site

as the human bones came from," he says. This bone yielded only horse DNA, he notes, whereas it would also have yielded human DNA if contamination had been a problem.

But such controls cannot definitively rule out contamination of the Cro-Magnon sample. Researchers such as Pääbo and Cooper see no way to resolve the issue, and are now increasingly questioning the value of studies such as Bertorelle's. "We know that contamination is almost impossible to avoid," laments Cooper. ■



Skull school: researchers in Italy have sparked a row over DNA studies of human evolution.

G. BERTORELLE

Experts blast US decision to back nuclear bunker-busters

Deep-sea sub aims to get to the bottom of a muddy issue

Quirin Schiermeier, Munich

Europe's most advanced ocean-research robot set out for an 11-week Arctic expedition last week, its most ambitious to date. Researchers hope that Victor 6000, one of the few craft that can reach depths of six kilometres, will provide insights into deep-sea corals and giant carbonate mounds — mud formations that can reach hundreds of metres in height.

"It's a fantastic tool," says cruise leader Michael Klages, a marine biologist at the Alfred Wegener Institute (AWI) of Polar and Marine Research in Bremerhaven, Germany. "This is the first time that we will be able to fully exploit its potential."

The €15-million (US\$17.8-million) craft, developed by IFREMER, France's national marine research agency, requires some 90 tonnes of winches, cables and navigation equipment, and a 3,000-volt generator. Data and control commands are relayed to and from the mother ship, the *Polarstern*, by a fibre-optic cable. Video cameras and robot arms can be used to examine organisms such as tube worms and sponges.

The first stop on the cruise, which is focused on areas of Europe's continental shelf, will be the Porcupine Seabight, a deepwater basin southwest of Ireland, where Victor 6000 will survey coral reefs and carbonate mounds. The *Polarstern* will then set off next month for the Håkon Mosby Mud Volcano, northwest of Norway.

The volcano releases methane-rich muddy fluids, which feed huge white mats of methane-dependent bacteria. Researchers plan to use Victor 6000 to probe the sediments and to survey the extent of the bacteria-covered slopes. Among other things, this should reveal how much methane escapes from the system and dissolves into the ocean.

The cruise will finally take the team to the AWI's 'House Garden', an established research area off Spitsbergen, an island group 650 km north of Norway. One study aims to deduce how benthic engineering — natural ploughing of the sediment by snails and crustaceans — creates suitable habitats for bacteria and other microorganisms.

"The cruise is an exciting project", says Jon Copley, a deep-sea biologist at the Southampton Oceanography Centre, UK. "Whenever someone visits the deep ocean with new technology, they almost always advance our understanding of what goes on out there."

♦ www.polarstern-victor.de



The United States plans to use nuclear warheads to beef up bunker-destroying bombs such as these.

Geoff Brumfiel, Washington

The US Congress has voted to plough \$15 million into developing Earth-penetrating nuclear weapons to destroy underground bunkers. But weapons experts have cast doubt on the scheme, arguing that this is an inappropriate use of nuclear technology.

The plan, which was proposed last year (see *Nature* 415, 945–946; 2002), aims to develop the weapons primarily to target hidden stores of biological or chemical weapons. According to the author of a new study, however, such agents may simply be dispersed by the weapons, and conventional weapons are better suited to the job.

President George W. Bush's administration welcomed last week's votes in both the House of Representatives and the Senate. General Richard Myers, Chairman of the Joint Chiefs of Staff, says that the radiation emitted by nuclear explosions could sterilize biological agents, and that the heat generated will destroy chemical weapons. Conventional weapons, he argues, are unsuitable as they are liable to spread such biological or chemical agents, creating a greater hazard.

But Robert Nelson, an astrophysicist and senior fellow at the Council on Foreign Relations, a think-tank based in New York, disagrees. In a paper shortly to appear in the journal *Science and Global Security*, he calculates the impact of an underground detonation of a nuclear device.

Nelson finds that rock or concrete surrounding a bomb would absorb heat and radiation, but would transmit the massive

shockwave caused by the explosion. A nuclear device equivalent to 10,000 tonnes of TNT, for example, would create a crater 200 metres wide, but only material within 11 metres of the centre would be sterilized. Large amounts of material would be ejected from the centre of the blast, along with radioactive dust created in the explosion. Nelson suggests that it would be better to use conventional weapons to block the entrances and ventilation shafts of such bunkers until they could be secured by friendly forces.

Nelson admits that his analysis is basic, but argues that the results are clear-cut. Sidney Drell, deputy director emeritus of the Stanford Linear Accelerator Center in California and an architect of the US stockpile-stewardship programme, designed to maintain nuclear weapons without testing, agrees that the study is sound. He believes that the crater and sterilization area could be marginally larger than Nelson estimates, but echoes Nelson's argument that such a blast might only act to disperse chemical or biological agents.

Drell adds, however, that he is more concerned by the \$6 million allocated last week by the Congress for research into another new type of nuclear device — low-yield nuclear weapons known as mini-nukes. The United States has had a self-imposed ban on such weapons since 1993, but last week's vote marks an end to this ruling. Drell is worried that the move blurs the line between nuclear and conventional weapons, ultimately increasing the likelihood that nuclear weapons will be used in anger.

NASA aims high with orbital transport system

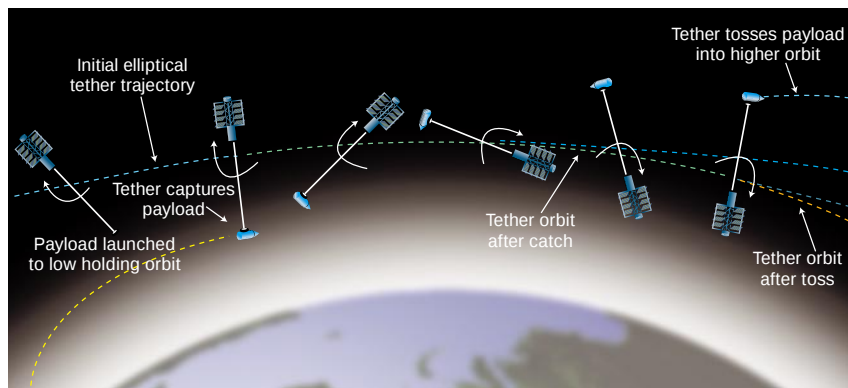
Tony Reichhardt, Washington

If it works, it could revolutionize space transportation. So NASA last week made a small but significant investment in a controversial technique for moving satellites around in orbit using slingshot-like tethers.

The technology works in two stages. First, satellites are boosted into higher orbits using a tether, perhaps 100 km or longer, that is rotating around an orbiting mass. A grapple mechanism at one end of the tether captures the satellite at the low point of the tether's rotation, then flings it into a higher orbit (see diagram).

The transfer of momentum from the tether to the satellite leaves the tether in a lower orbit. So the second stage of the process uses a power source on the central mass to run an electric current through the tether. Once the current is flowing, the tether experiences a force that results from its movement through Earth's magnetic field. This pushes it back into its original orbit.

NASA doesn't expect to see this kind of tether system demonstrated before 2010, but has awarded \$4 million to four teams work-



Swinging time: a rotating tether could be used to grab a satellite and toss it into a higher orbit.

ing on it, as part of a programme designed to foster innovative propulsion concepts. The technology, known as Momentum-Exchange/Electrodynamic-Reboost, will be evaluated over two years using computer simulations and other basic research.

The idea of tether propulsion is an old one but it has yet to prove itself, says Enrico Lorenzini, head of a space-tethers research

group at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, who has received one of the NASA grants. NASA, the US Department of Defense, student researchers and small companies have all flown tether experiments in space with limited success. A pair of US-Italian tethers that were unfurled from the space shuttle in 1992 and 1996 proved that tethers could produce current when dragged through the Earth's upper atmosphere, but hardware problems cut both experiments short.

A NASA-funded tether experiment, known as the Propulsive Small Expendable Deployer System or ProSEDS, was to have been launched this spring. The system would have been deployed from the spent stage of the Delta rocket that launched it. By interacting with the Earth's magnetic field, the tether would have dragged the rocket back and released it into the atmosphere, where it would have burned up.

The experiment was intended to show that tethers can be used to clear space junk without using rocket fuel. But after the accident involving the space shuttle Columbia, NASA rescheduled the launch for next February. The agency was worried that if something went wrong, the space station would have to use precious manoeuvring fuel to avoid a wayward tether.

Lorenzini and other researchers in the field see the mission as a key milestone in winning confidence that tethers are safe and practical. If the experiment is successful, they believe that other applications will follow. Pioneers in the field say that last week's grant also shows a new level of space-agency support for more advanced tether-propulsion ideas.

Robert Hoyt of Tethers Unlimited in Lynnwood, Washington, is another of the researchers who has been given a grant by NASA. "It's the first time what I would call real NASA money has been spent on the concept," he says.

Legal row looms for gene-map firm

Rex Dalton, San Diego

A leading company in the International HapMap Project, a collaboration aiming to map genetic variation in the human genome, is embroiled in a high-stakes legal battle.

At \$120 million, HapMap is one of the largest genetics projects since the sequencing of the human genome. Researchers plan to use five separate technologies to identify patterns of genetic variation that affect disease and health (see *Nature* 412, 105; 2001). Microarray technology developed by Illumina of San Diego, California, will be used in the United States, Britain and Canada to produce about half of the HapMap data.

But Illumina faces a legal demand from biotechnology giant Applera, based in Norwalk, Connecticut, for \$30 million in damages — slightly more than Illumina's operating expenses for last year.

Applera's subsidiary company Applied Biosystems in Foster City, California, invested \$10 million in Illumina in 1999 as part of a plan to develop new genotyping technologies. Illumina says that Applera failed to fulfil its contractual obligations, such as providing reagents, and over the past year both firms have been developing their own technologies.

Last December, Applera sought to resolve the dispute through arbitration. Fearing that it could lose the rights to continue work on its core technology, Illumina filed a suit against

Applera alleging that the arbitration was intended to put it out of business. The arbitration was blocked by a state court in San Diego in February, but Applera has since filed a counter-suit. Illumina says that Applera is demanding \$30 million in return for work on the collaboration and its initial investment.

Illumina has informed its shareholders that the litigation could have an "adverse affect on its business, financial conditions and results of operations" if it is unable to defend itself successfully.

Timothy Kish, Illumina's vice-president and chief financial officer, told *Nature* that the legal action will not stop the company supplying its microarray technology to HapMap scientists. But lawyers for the company said the case could have a severe impact on the HapMap project.

Applera officials declined to be interviewed, but issued a statement saying that the firm's intention was to protect its investment, not to interfere with scientific research.

Researchers on HapMap projects in the United States and Britain say they are only vaguely aware of the litigation. "Potentially, we are concerned," says David Bentley, a geneticist at the Wellcome Trust Sanger Institute near Cambridge, UK, although he adds that he is confident that Illumina will be able to honour its agreements.

Biologists moot daunting plan to track viruses

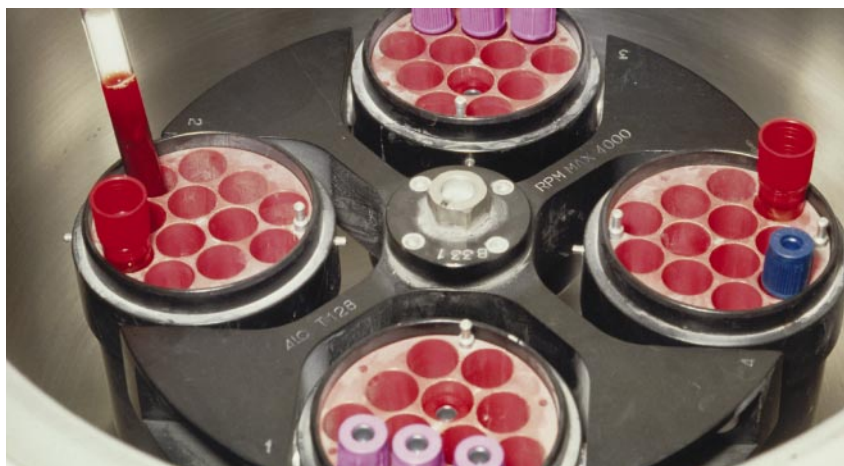
Jonathan Knight, San Francisco,
and Alison Abbott, Munich

Three US researchers have proposed an ambitious project to sequence the genetic code of every virus known to infect humans. They plan to use the information in a database to track new and emerging viral infections. But critics say that it is not always possible to tell harmful and benign viruses apart, and that such a system could generate numerous false alarms.

The proposal, to be published in the July issue of *Emerging Infectious Diseases*, would involve collecting viral particles from blood plasma from a variety of sources, including the Red Cross and hospital labs. The genetic material would then be extracted and sequenced, and a database built up.

Methods for using centrifugation to purify viruses from large volumes of plasma were developed in the 1960s by the proposal's principal author, Norman Anderson, president of the Viral Defense Foundation based in Kensington, Maryland. Once fully automated, the sequencing process should be rapid enough to spot new viruses within days of obtaining blood samples, he says.

Anderson developed the proposal with his son Leigh, who is chief executive of the Plasma Proteome Institute, a Washington-based organization that promotes protein research, and John Gerin, director of the molecular virology and immunology division at Georgetown University Medical Center in Rockville, Maryland. Although the plan is in its early stages and has not yet been costed, the sequencing of viruses en masse



In a spin: centrifuges will be key for separating viral particles from blood samples for genetic analysis.

has already been achieved in marine biology. A team at San Diego State University, which is collaborating with Anderson, last year sequenced hundreds of viruses from the Pacific Ocean (M. Breitbart *et al. Proc. Natl Acad. Sci. USA* **99**, 14250–14255; 2002).

Anderson's scheme will face additional technological challenges. The marine viruses had DNA genomes, but the genomes of most human viruses are made up of RNA. This adds an extra step to the sequencing process, as the RNA must be converted into DNA.

Even if the project is feasible, others say that the real challenge will be interpreting the information. "You will end up with a lot of results that cannot be linked to disease," says Christian Drosten, a virologist at the Bernhard

Nocht Institute for Tropical Medicine in Hamburg, Germany. The people from whom the blood was taken would have to be screened to see if their immune systems were reacting to a new virus, which would slow the tracking process considerably, says Drosten. Spotting new viruses in blood could also lead to unnecessary blood-supply scares, he adds.

Anderson says that such concerns can be overcome. He believes the danger of new viruses could be assessed by monitoring epidemiological reports from the regions in which they emerge. The system could also help to solve a number of disease puzzles, he adds, as the database would allow researchers to spot correlations between viruses and diseases that are not completely understood. ■

Threat of closure hangs heavy over primate centre

Rex Dalton, San Diego

The future of the Duke University Primate Center in Durham, North Carolina, considered by evolutionary biologists to be



Duke University's primate centre is seen by experts as a unique resource for studying lemurs.

unique and irreplaceable, is hanging in the balance. Administrators at Duke University could decide within a month not to seek a new permanent director and to phase the animal centre out of existence.

The decision will come at the end of an 18-month review period triggered by criticism that the facility had too many animals but carried out too little research. Duke officials sought to reorient the centre towards teaching and research, rather than conservation, and Duke officials say that it has improved. The colony has fallen from a peak of more than 600 animals to about 250, following the introduction of tight reproductive controls and a move to donate excess animals to zoos.

The Duke centre is home to the world's primary research colony of living prosimians, a branch of primates that evolved 50 million years before apes and monkeys. Often dubbed 'living fossils', the

centre's prosimians, mostly lemurs, provide researchers with an unmatched resource for studying aspects of early primate evolution, from blood chemistry to morphology. "The centre is a gem beyond compare," says Anne Yoder, an evolutionary biologist at Yale University in New Haven, Connecticut.

Elwyn Simons, a palaeontologist at Duke, says that about 50 projects studying living lemurs are under way, with about 40 projects involving the facility's prosimian fossil collection. "We are very impressed with the ramping up of research," says Jim Siedow, Duke's vice-provost for research.

The facility, which was created in the 1960s and has an annual operating budget of about \$1.2 million, could probably not be duplicated because some prosimian species are now endangered. Lemurs live in the wild only in Madagascar, where they are on the brink of extinction. ■

Setback forces rethink for cell signalling alliance

Pasadena Cells can be uncooperative workmates, as scientists in the Alliance for Cellular Signaling (AfCS) have found to their cost. The AfCS, a multi-lab collaboration, was launched last year to map the signalling networks in cells using a range of techniques (see *Nature* **420**, 600–601; 2002). Two cell types — B lymphocytes and heart muscle cells — were to be used as the first models.

But researchers at the third AfCS meeting, held last week in Pasadena, California, reported that both cell types have proved to be incompatible with one of the alliance's most important tools: the RNA-interference technique that allows scientists to knock out specific proteins in a signalling pathway. "We don't know exactly why the primary cells are misbehaving," said Ron Taussig, associate director of the AfCS at the University of Texas Southwestern Medical Center in Dallas.

Scientists at the meeting said they now plan to use a macrophage cell line, which is suitable for RNA-interference studies and function normally in culture. "We may have lost a little time," says Taussig, "but not too much."

Diagnostic drive aims to combat killer diseases

Geneva The World Health Organization (WHO) and the Bill & Melinda Gates Foundation have joined forces to set up a new organization to promote the search for cheap, easy-to-use diagnostic tests for infectious diseases, particularly in the developing world.

The Foundation for Innovative New Diagnostics (FIND) will support companies and academic institutes in creating and validating new diagnostic methods that use



Labour of lab: medical staff in Africa still rely on inefficient technology to diagnose viral diseases.

simple methodologies, such as dipsticks, rather than sophisticated equipment.

FIND has an initial budget of \$30 million for the next five years, and will call for proposals for tuberculosis tests in the coming weeks. "Tuberculosis is a major killer" in the developing world, says Charles Morel, director of the WHO's tropical-medicines programme. "Yet we are relying on very old-fashioned and inefficient diagnostic methodologies with which local medical support cannot cope."

Bill promises genetic privacy for US citizens

San Francisco Discrimination on the basis of genes is about to become illegal in the United States. After being held up for six years by wrangling over the details, a bill that sets new genetic-privacy standards has cleared a key Senate committee, and is now expected to pass without further challenge. President George W. Bush has said he will sign the bill.

Genetic testing can help doctors to estimate a patient's susceptibility to a variety of diseases, including breast cancer and Parkinson's disease. But many people refuse the tests amid fears that a positive result could see them forced to pay higher insurance premiums or even lose their jobs. The new bill bans such discrimination.

Employers and insurance companies had

objected to a blanket ban on the use of genetic information. The compromise reached on 20 May allows insurers to collect genetic data on patients with genetic disorders for the purpose of managing their care. Employers will be able to use DNA testing to check for adverse effects from exposure to workplace chemicals or radiation.

Taiwan slams health officials for 'lack of help' over SARS

Taiwan Officials in Taiwan, which now has more cases of severe acute respiratory syndrome (SARS) than anywhere except China and Hong Kong, have accused the World Health Organization (WHO) of failing to help control the outbreak there. Taiwan has been excluded from WHO meetings and has had difficulty in getting diagnostic and other materials to fight the disease (see *Nature* **422**, 652; 2002), as a consequence of its long-standing diplomatic dispute with China.

Taiwanese president Chen Shui-bian has led the push for observer status at the WHO, but on 19 May the organization's ruling body, at the beginning of its ten-day meeting in Geneva, voted to reject the request.

The WHO blames Taiwan's lack of a centralized disease-control policy for the extent of its SARS outbreak. The organization sent two officials to Taiwan in early May — the first time that WHO representatives have visited Taiwan in decades.

The number of new SARS cases elsewhere in the world is now falling rapidly. Some researchers have expressed an interest in establishing collaborations in Taiwan to test new diagnostic devices on the large number of suspected cases there.

Flame rekindled for Alzheimer's vaccine

Zurich Hopes were revived last week that a vaccine could slow the deterioration seen in patients with Alzheimer's disease, after a group at the University of Zurich reported some improvement in vaccine-treated patients.

The vaccine, produced by the Irish firm Elan, generates an immune response against the insoluble β -amyloid plaques found in the brains of Alzheimer's patients. But clinical trials of the vaccine were stopped more than a year ago when 6% of the subjects developed brain inflammation.

The Swiss scientists monitored 30 patients who had taken part in the abandoned trial, and found that 20 of them did not deteriorate as expected. Some patients even performed better in memory tests than before they were immunized. Moreover, antibodies against β -amyloid were found in the blood of the 20 patients, but not in that of the other 10 (C. Hock *et al. Neuron* **38**, 1–20; 2003).

Analyses at Elan have not found the same benefit in the entire set of patients.

Record-breaking bloom kicks up a stink

Munich A titan arum at the Botanical Gardens in Bonn, Germany, last week produced the largest flower ever measured. The bloom reached 2.7 metres — 8 cm more than the previous record, set 70 years ago.

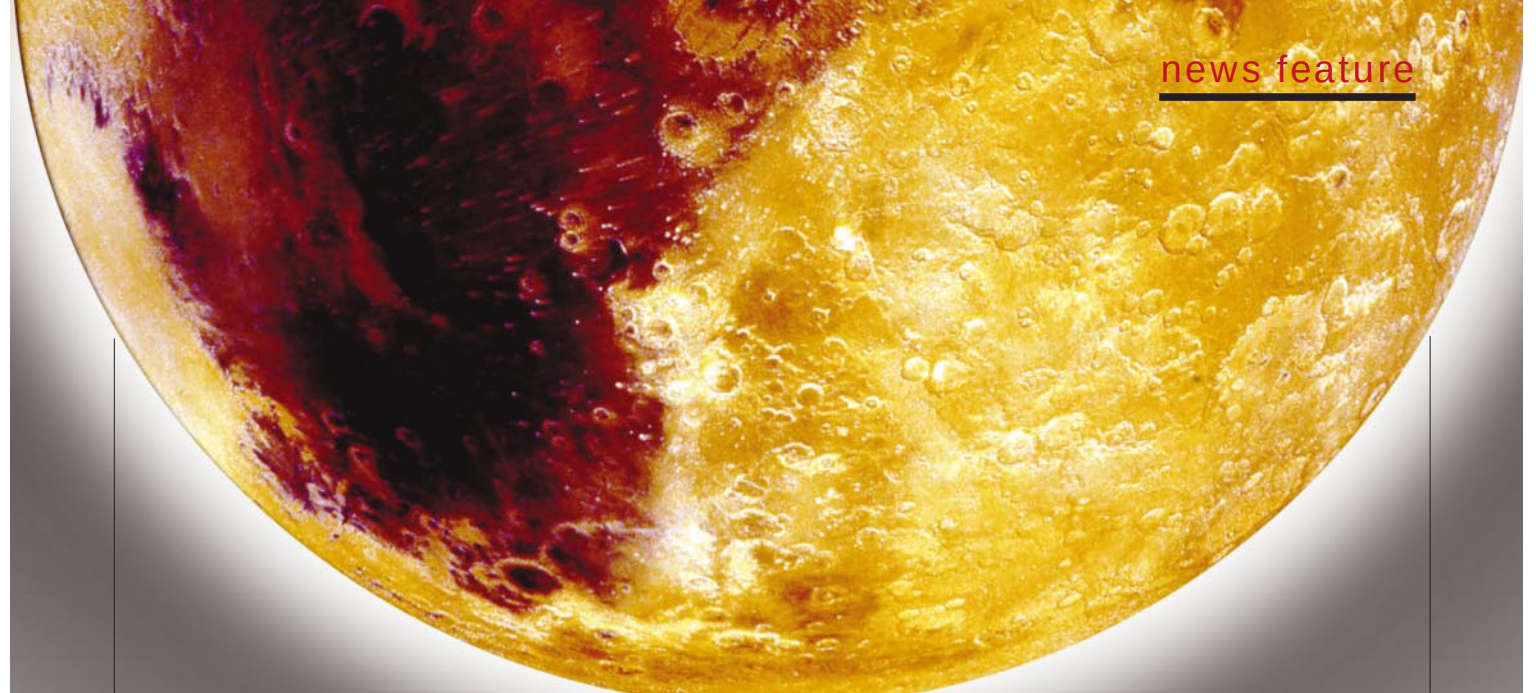
The titan arum, or *Amorphophallus titanum*, is indigenous to the equatorial rainforests of Sumatra in Indonesia. The Florentine botanist Odoardo Beccari first discovered the plant in 1878, and sent seeds to Kew Royal Botanical Gardens in London, where it first bloomed in cultivation in 1889.

In its shape, brownish colour and pungent smell of rotten flesh, the titan arum mimics a carcass in decay, thus attracting insects which deposit their eggs inside the flower, pollinating the plant in the process.

The rare bloom lasted only a matter of days, after which the flower's spadix — its fleshy central column — collapsed under the weight of the blossom.

► www.botanik.uni-bonn.de/botgart/amorpho2003e.html





Mars attracts!

Interest in the red planet is about to peak, as three missions prepare to join the hunt for water and life on one of our closest neighbours.

Astrology is rightly maligned, but the position of the planets has at least some predictive power. Every 26 months, Mars and Earth come into an alignment that minimizes the fuel needed to journey between the two. And every time that happens, Mars researchers find themselves excited, nervous and overworked, as they prepare to launch spacecraft in which they have invested years of effort.

That alignment is upon us again, and three new craft are scheduled to take off next month. In the pages that follow, *Nature* introduces those missions in graphical form (see page 474), and reveals the human stories behind the hardware.

If some involved feel worried, it is partly because they have previously worked on craft that launched but never made it to Mars (see page 477). But such failures have not deterred the world's space agencies. The Japanese probe Nozomi is already en route to Mars, and many other missions are on the drawing board to take advantage of future favourable planetary alignments.

Some planetary scientists feel that the degree of attention lavished on Mars is unwarranted. Mars is already the best-mapped planet after the Earth, and we know more about its climate and geology than that of any of our other neighbours. Yet the public, funding agencies and scientists are still drawn to the red

planet, for the simple reason that extraterrestrial life may once have existed there.

In the case of NASA's two new Mars Exploration Rovers, the search for life will be conducted through a proxy search for water. Liquid water is not thought to exist on Mars today. But water would have been necessary if life were present in the past, so the rovers will hunt for its geological calling card.

At the dawn of each martian day, the rovers will receive instructions based on images sent back to the Earth by their four sets of onboard cameras. They will be told what site to visit and what to do there, but not how to get there — signals take around ten minutes to travel between Mars and Earth, so onboard navigation software will be used to avoid obstacles in the rovers' paths. Once a target site has been reached, numerous instruments can be deployed: grinders will drill into the rock, for example, and magnets will be used to collect particles for further analysis onboard the rover.

The Beagle 2 lander, part of the Mars Express mission run by the European Space Agency (ESA), will be stationary, but will make up for its sedentary nature by deploying some innovative equipment (see page 476). One example is the 'mole', a bicycle-pump-shaped device that will crawl along the planet's surface and can be directed vertically into the soil by retracting and unleashing a pair of spring-

loaded weights. The device will collect samples in a cavity at its head; winding in the mole's power cable allows the samples to be retrieved. A mass spectrometer on Beagle 2 will then reveal, amongst other things, the relative amounts of different carbon isotopes in the rock. Living organisms, such as bacteria, take up these isotopes in different ways, so the analysis could reveal whether life was present when the rocks formed.

Both the NASA and ESA landers will have relatively short missions — around three and six months, respectively. But in the sky above, two orbiting craft will continue to study Mars, probably until 2006. Mars Express will probe the composition of the martian atmosphere and surface. Nozomi, which will arrive four years later than planned because a malfunctioning thruster forced engineers to adjust its route, will study how the planet is buffeted by the solar wind, the stream of charged particles emitted by the Sun.

Such mishaps as befell Nozomi are common in planetary exploration — and so the nerves of ESA and NASA scientists are likely to be frayed over the next couple of months. Once the launches are over, the respite will only be temporary. Expect high levels of anxiety and edginess among Mars researchers early next year, as the crafts on which they have staked so much approach their final destination. ■

Jim Giles, associate news and features editor.

Released into the atmosphere, a probe falls towards the surface at over half a kilometre per second.

Fantastic journeys

Early next year, three instrument-laden landers will touch down on Mars. The Japanese craft Nozomi will enter orbit around the planet shortly afterwards. By spring 2004, data on the red planet will be flowing thick and fast.



Three hundred metres above the surface, the lander is dropped from the probe. A cluster of airbags inflates around the lander and a parachute slows its descent.

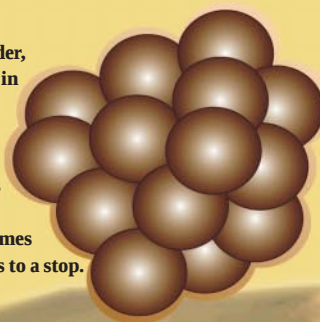
Meridiani Planum

One of NASA's twin Mars Exploration Rovers will touch down at Meridiani Planum, where it will find itself on the opposite side of Mars to its sibling craft. Grey deposits of the mineral haematite attracted NASA to this site. Haematite — an oxidized form of iron — rarely forms on Earth in the absence of water. Its presence at Meridiani suggests that hot springs may once have gurgled through the rocks there. Using a high-power microscope, the rover will examine the size and the orientation of haematite grains, which should provide clues about how the mineral formed.

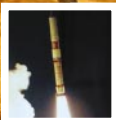
Gusev Crater

NASA staff plumped for this site after noticing its resemblance to lake beds seen on Earth. Sediment seems to have accumulated at the point at which a 900-km-long valley courses into the crater. One of NASA's rovers will land in the deepest part of the crater, and use an abrasion tool to grind away the surface layers of sedimentary rocks. Measuring the X-rays and α -particles emitted by the lower layers of rock will provide data on their composition and the conditions under which they formed — including whether water was present at that time.

The lander, encased in airbags, hits the ground, bounces about a dozen times and rolls to a stop.



4 July 1998



Japan launches Nozomi from the Kagoshima Space Center

2 June 2003



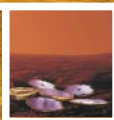
Europe's Mars Express due for lift-off from Baikonur Cosmodrome, Kazakhstan

5 & 25 June 2003



NASA's rovers set for launch from Cape Canaveral, Florida

25 December 2003



Mars Express scheduled to release Beagle 2

January 2004

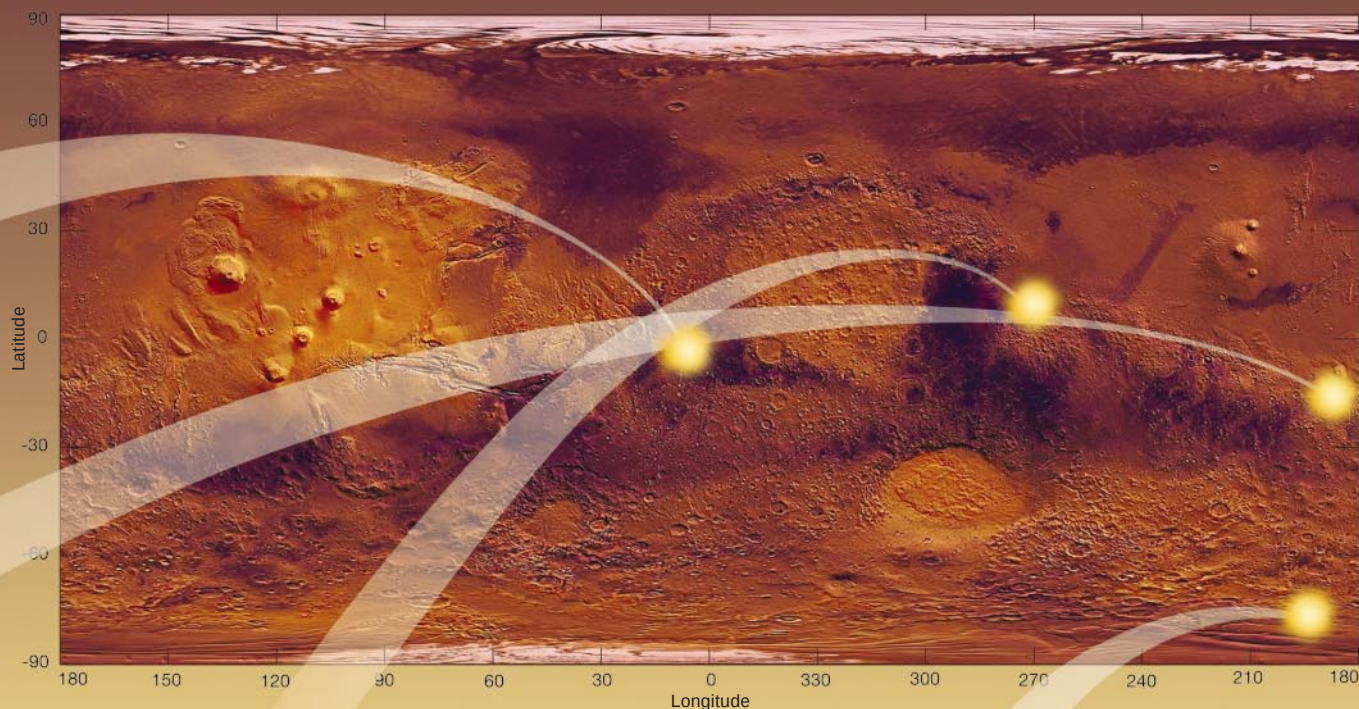


NASA's rovers due to touch down on Mars

Early 2004



Nozomi expected to enter orbit around Mars



Isidis Planitia

Europe's Beagle 2 will land in a boundary zone. The Isidis Planitia, a crater 1,600 km wide, separates the red planet's more cratered southern hemisphere from the flatter northern half. The flat, smooth features of Isidis Planitia should make it a safe landing site, and its volcanic domes and small channels may have preserved evidence of life. Beagle 2 carries a set of ovens, which will heat promising rock samples at a variety of temperatures to determine the nature of any carbon-containing matter.

Polar Lander

The fate of NASA's Mars Polar Lander isn't known for certain, but officials assume that fragments of the lander litter a depression near the planet's southern pole. Contact with the craft was lost on 3 December 1999, as it began the final stages of its descent to the martian surface. The lander would have touched down here and studied minerals near the planet's poles, using a drill to bore into the surface in search of water. NASA believes that a jolt to the lander's leg shut off the landing engines, causing it to crash with the loss of all instruments.

The solar-panel petals unfold, the lander surveys its surroundings and begins exploring. NASA's Mars Exploration Rover is shown here, but the European Space Agency's Beagle 2 will also use a parachute and airbag to land safely on Mars.

For an extended interactive version of this graphic, visit our Mars web focus at www.nature.com/nature/focus/mars

Research: Hannah Hoag

Art: Cliff Saunders and Ann Thomson

Photos: NASA, ESA, ISAS, Beagle 2



Are you on board?

The Beagle 2 Mars lander has had some unusual backers — among them British pop stars and artists. Declan Butler finds out how one researcher's publicity drive got the project off the ground.

For a man with a reputation as a political animal, Colin Pillinger's appearance might seem rather incongruous. The planetary scientist, based at the Open University in Milton Keynes, UK, speaks in a rural drawl and opts for a leather jacket and checked shirts rather than sharp suits. Scarecrow straggly hair and muttonchop whiskers complete a look that might be said to be more in keeping with his sideline in farming than with his role in publicizing a multi-million-pound space mission.

But Pillinger's image is deceptive. When the Mars Express mission, under the auspices of the European Space Agency (ESA), blasts off on 2 June, it will carry the Beagle 2 lander — a device whose existence justifies Pillinger's claim to be a "professor of PR". If it hadn't been for Pillinger's ability to generate publicity, and his belief in his work, ESA's mission could have been going to Mars without a landing craft.

The political manoeuvring behind Beagle 2 dates back to 1997, when ESA decided to develop a craft that would travel to Mars and study the planet from orbit. Like other ESA spacecraft, the agency planned to fund Mars Express and its launch, whereas member nations were to design and pay for the instruments it carried.

Pillinger, who was renowned for his studies of martian meteorites, wanted to include a lander that would search for evidence of past or present life by analysing rock, soil and atmospheric samples. The plan had the backing of ESA, but the Particle Physics and Astronomy Research Council (PPARC), the main UK funding body for such projects, gave the idea a cool reception. Officials say that Beagle 2's budget was initially poorly defined, and that the council's space funding was already committed elsewhere.

But Pillinger knew what buttons to press.



Colin Pillinger (above) has boosted the profile of his Beagle 2 project by collaborating with figures such as artist Damien Hirst, whose dot painting (right) will calibrate the lander's camera.



It was originally claimed that the mission would cost about £25 million (US\$41 million) — a snip for such an ambitious project, and that sponsorship and advertising would provide two-thirds of the cost. M&C Saatchi, a London-based advertising firm, was signed up to search for sponsors and even sell the name of the mission.

Media interest moved up a notch when Blur, a successful British pop band, agreed to compose a nine-note tune that the lander will beam back to Earth to signal its safe arrival. Damien Hirst, an artist famous for suspending animals in tanks of formaldehyde, provided a further publicity boost when he designed the painting that Beagle 2 will use to calibrate its onboard camera.

The strategy paid off. In August 1999, the UK government agreed to plough £5 million into Beagle 2. About 20 companies, including the McLaren Formula One team, also offered to help. And after being impressed by Pillinger's formal science proposal, the PPARC pitched in £5 million.

The resulting lander has attracted much acclaim. At 65 kilograms and 0.95 metres across, the ratio of instruments to mass is unprecedented, says Steven Squyres, an astronomer at Cornell University in Ithaca, New York, and part of the team behind NASA's Mars rovers (see opposite page). "To put together a package that capable with

To put together a package that capable with the challenges they faced is just incredible. Steven Squyres

the technical and financial challenges they faced is just incredible," he adds.

But Beagle 2's financing did not turn out quite as planned. Industrial companies gave help in kind, but the larger aerospace firms worked for free only at the start of the project. M&C Saatchi has ended its involvement following the departure of Matthew Patten, head of the company's sponsorship arm and the project's internal champion.

The Beagle 2 team decline to discuss the costs involved, saying that some sponsors wanted to remain anonymous. But *Nature* has learned that costs for the mission may stand at around £40 million — about £15 million above original estimates. The shortfall has been made up by the government and ESA, although two-thirds of the £17 million contributed by ESA must be repaid by the British government.

Pillinger refused to discuss this and other matters with *Nature*, citing a disagreement over some correspondence from him that was published in the magazine in December 1996. Some of those who have worked with him say that he can be cantankerous. According to one astronomer, he can be "astoundingly tactless", and is liable to call people who he thinks are wrong "bloody fools".

Yet space researchers are ultimately full of admiration for Pillinger, and for the way in which he has overcome a lack of initial interest in his Mars project. "Something like Beagle 2 doesn't happen without someone with the vision and drive to force it through," says Squyres. And Pillinger could yet find that the budget is no longer an issue. If the lander finds evidence of life on Mars, sponsorship deals should be much easier to strike in the future.

Declan Butler is *Nature's* European correspondent.
♦ www.beagle2.com

The comeback kids

NASA's Jet Propulsion Laboratory practically invented planetary exploration. Then, in 1999, it lost two craft in quick succession. Tony Reichhardt meets the staff behind two new Mars rovers, which could restore the lab's reputation.

Those working on this year's Mars missions probably want to forget the precedent set by previous attempts. More than half of the 31 craft sent to Mars since 1960 have failed, including six of the last ten. "This is a very unforgiving business," says Richard Cook, an engineer at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California. "We're still in the phase where landing on Mars is a miracle every time."

Cook knows just how unpredictable planetary exploration can be. In 1997, at the age of 32, he was flight operations manager for Mars Pathfinder — the first craft to land on the red planet in two decades. Two years later, he was project manager for the Mars Climate Orbiter and the Mars Polar Lander. The first of these veered off course because of a mix-up over imperial and metric units; a software error sent the second crashing into the planet. After contact with the lander was lost, Cook's glum face appeared on television every night, waiting for a signal that never came.

After the losses, the staff of a lab that had built pioneering spacecraft that journeyed to the edge of the Solar System were cast as a bunch of incompetents. "It caused a lot of soul-searching," says Steve Squyres, a planetary scientist at Cornell University in Ithaca, New York. But NASA kept faith, choosing JPL and its staff, including Cook, to develop its two Mars Exploration Rovers. As launch dates for the rovers approach — 5 June and 25 June provisionally — morale at the lab is high, even if the pressure is on. "This is the best team that I have ever seen at any JPL project," says Squyres, who is principal investigator for the rovers' science package. "We all feel like we must succeed."

Staff at the lab say that the challenge of the new mission has prevented them from dwelling on the past. Joy Crisp, project scientist for the rovers, has indirect knowledge of the pain caused by the lost missions — her husband, David Crisp, was a project scientist for the Mars Polar Lander. "The complexity of these new missions is amazing," she says.

This time there will be two rovers instead of one, each travelling many times the 100 metres that Pathfinder ventured from its landing site. Improved scientific instruments will also allow more detailed studies of rock chemistry and mineralogy. Crisp coordinates the roughly 100 scientific staff who will use these instruments. They will meet twice daily to weigh up their research objectives and sta-



tus reports before deciding what to do next.

JPL staff have also been kept busy planning how to deal with working in martian time. Rover operations start at the crack of the martian dawn, but a day on Mars is 24 hours and 37 minutes long. Shifts and meeting times at JPL have to move forwards by 37 minutes each day to stay in sync, and staff have to try to adjust their lives accordingly. Project managers worry about 'martian jet lag', and have brought in experts on fatigue to advise them on how to arrange work shifts to avoid burn-out. Crisp lived on martian time for the 30 days of the Pathfinder mission. "That's about as much as you'd want to do," she says.

Matters might be better if the team were heading into the landing well-rested. But Squyres says that the project's timescale has been "absolutely brutal". As well as worrying about every last bolt and wire on the spacecraft, the team has rehearsed whatever operations it can ahead of time. Four times,

A day on Mars lasts 24 hours and 37 minutes, so project managers are worried about 'martian jet-lag'.

Ready for action: the team behind NASA's rover mission to Mars includes Joy Crisp and Richard Cook (insets), and Steve Squyres, seen here testing a model rover in the desert.

scientists have conducted practice sessions with mock rovers in the deserts of the American southwest. By orchestrating its procedures before the rovers reach Mars, the JPL team hopes that it won't waste precious time debating which rock to examine next, how best to drill into it, or what images to take.

The practice runs may have been tiring, but Crisp says that they feel much more prepared this time. Even though Pathfinder was a success, she says, "we felt like we were winging it" after it landed on the martian surface. That might have been acceptable for Pathfinder, which was primarily intended as a technology demonstrator. But it won't be acceptable for the rovers, which need to do more than just touch down in one piece. "If all we do is land," says Squyres, "we don't score any points at all."

And so the researchers head towards the launch pad acknowledging the pressure, but not allowing themselves to be overwhelmed by it. Cook is philosophical. "The line between success and failure is very thin," he says. "You are literally dependent on 10,000 things going exactly right, and if you're wrong on one of them, that can be it."

Tony Reichhardt writes for *Nature* from Washington.

R. SULLIVAN; STOCKTREK/CORBIS (BACKGROUND)

Challenging the tyranny of impact factors

A recent Commentary aroused a lively debate. In this issue we publish some responses.

Sir — Peter A. Lawrence's Commentary "The politics of publication" (*Nature* **422**, 259–261; 2003), about journal mania and the tyranny of impact factors, was like a breath of fresh air. How did this self-inflicted misery arise, and how on Earth are we going to get rid of it?

Clearly it arises largely from laziness and poor homework on the part of senior scientists and bureaucrats (the dividing line is sometimes thin). It is amazing how unscientific some of them can be once outside their own narrow field. Eugene Garfield, who invented the wretched impact factor, himself said that it is not appropriate for ranking individuals (see [www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html)).

Astonishingly, these facts are not known (or are ignored) by some selection committees. Serious studies have been done, such as that of P. O. Seglen (*Br. Med. J.* **314**, 498–502; 1997), which shows that the citation rate for individual papers is essentially uncorrelated to the impact factor of the journal in which it was published. This happens because of the very skewed distribution of citation rates, which means that high-impact journals get most of their citations from a few articles.

The distribution for *Nature* is shown in Fig. 1. Far from being gaussian, it is even more skewed than a geometric distribution; the mean number of citations is 114, but 69% of papers have fewer than the mean, and 24% have fewer than 30 citations. One paper has 2,364 citations but 35 have 10 or fewer. (The Institute of Scientific Information, ISI, is guilty of the unsound statistical practice of characterizing a distribution by its mean only, with no indication of its shape or even its spread.) ISI data for citations in 2001 of the 858 papers published in *Nature* in 1999 show that the 80 most-cited papers (16% of all papers) account for half of all the citations.

In my own work, for example, I have published a *Nature* (impact factor 27.9) article with only 57 citations, and an article in *Philosophical Transactions of the Royal Society* (impact factor 3.1) with more than 400 citations. Gratifyingly, this represents roughly my own assessment of the relative worth of the articles, but it emphasizes the ludicrousness of the recent current obsession with journal impact factors, especially at a time when, because of the Web, it matters less than ever before where an article is published.

Perhaps one way to cope with the problem is to turn it on its head. Candidates can judge institutions by the questions

they ask, rather than the other way round. Any selection or promotion committee that asks you for impact factors is probably a second-rate organization. A good place will want to know about the quality of what you have written, not where you published it — and will be aware that the two things are uncorrelated. A useful method for job interviews that has been used in our department is to ask candidates to nominate their best three or four papers, then question them on the content of those papers. This selects against publication of over-condensed reports in high-impact journals (unless it is one of the relatively few genuinely important papers of this type). It also selects against 'salami slicing', and is a wonderful way to root out guest authors, another problem of the age. Experience has shown that candidates can have astonishingly little knowledge of the papers on which their names appear.

David Colquhoun

Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK

Editors are meant to be judges, not postmen

Sir — Congratulations to Peter A. Lawrence¹ for his forceful denunciation in Commentary of the pernicious but pervasive view that publication in the 'right' journals is the same as publication of the best science. In particular, scientific editors who serve largely as postmen forwarding referees' comments to authors and expecting total compliance with referees' requests, however ill-judged, should respond to his plea that they revert to their proper function as judges both of the manuscript offered and of the referees' views. In my experience, there are still some proper editors, most often to be found on those learned-society-based journals that properly support their editors.

Lawrence elected "the original paper by Michael Berridge and Robin Irvine on phosphoinositol and cell signalling", published in 1984, to show that the difficulties of publishing 'in the right place' are far from new. But in doing so he dived straight into the 'overcompression of information' trap that he rightly castigates. Rather than citing the first paper to establish the role of inositol 1,4,5-trisphosphate in Ca^{2+} signalling, he referred to a review that Berridge and Irvine wrote soon after. The original experimental paper², also in *Nature*, had

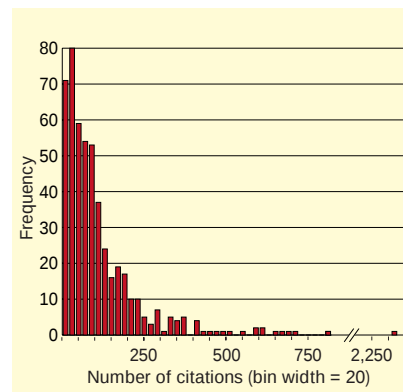


Figure 1 Distribution of the number of citations in five years for 500 biomedical papers published in *Nature*: 100 papers published in each of 1981, 1984, 1988, 1992 and 1996 were chosen at random, and for each paper the number of citations in the subsequent five years was counted. Data provided by Grant Lewison (Department of Information Science, City University, London EC1V 0HB, UK).

four authors, and was far from the 'original' paper on this topic. The entire field started in 1953 (ref. 3), and notions of the possible involvement of inositol lipids and phosphates in signalling reactions go back to the late 1960s (ref. 4).

My experiences in dealing with *Nature* during the decade leading up to ref. 2 were frustrating but sometimes, in the longer term, beneficial. Once the idea that phosphoinositide breakdown is implicated in cell signalling had grabbed us, we envisaged inositol 1:2-cyclic phosphate as a signal in stimulated cells in a paper in *Nature New Biology*⁵ — we were wrong. Two more steps were essential before Streb *et al.*² could home in on the Ca^{2+} -mobilizing action of inositol 1,4,5-trisphosphate. The first was to recognize a possibly causal relationship between receptor-driven phosphoinositide hydrolysis and Ca^{2+} mobilization⁶. Then we had to appreciate that receptor stimulation provokes $\text{PtdIns}(4,5)\text{P}_2$ hydrolysis^{7,8} rather than the hydrolysis of PtdIns that we had previously postulated⁶. It was the latter discovery that identified $\text{Ins}(1,4,5)\text{P}_3$ as the likely product of the receptor-regulated phospholipase.

Both of these advances were first submitted to *Nature* and were rejected without reaching referees. On each occasion I was in the fortunate position of having an immediate opportunity to expound the new idea in detail in an

invited review or symposium paper. As a result, the first impact of each paper may have been slower, but availability of a more detailed initial exposition than could have appeared in a 'top' journal probably helped the ideas to be understood more readily.

Bob Michell

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

1. Lawrence, P. A. *Nature* **422**, 259–261 (2003).
2. Streib, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67–69 (1983).
3. Hokin, M. R. & Hokin, L. E. *J. Biol. Chem.* **203**, 967–977 (1953).
4. Michell, R. H. in *Phosphoinositides and Receptor Mechanisms* (ed. Putney, J. W.) 1–24 (Alan R. Liss, New York, 1986).
5. Michell, R. H. & Lapetina, E. G. *Nature New Biol.* **240**, 258–260 (1972).
6. Michell, R. H. *Biochim. Biophys. Acta* **415**, 81–147 (1975).
7. Creba, J. A. et al. *Biochem. J.* **212**, 733–747 (1983).
8. Michell, R. H. *Phil. Trans. R. Soc. Lond. B* **296**, 123–138 (1981).

The system rewards a dishonest approach

Sir — I strongly support the views of Peter A. Lawrence in his Commentary about the politics of publication (*Nature* **422**, 259–261) and congratulate *Nature* for publishing this piece.

As Lawrence says, the assessment of science is, regrettably, moving towards an “audit society”, in which the fact of having published in high-impact journals is seen as more important than the content of the papers themselves.

There are two further problems that arise from this assessment of individuals through the locations of their publications. First, only a minority of submissions get sent out for review in high-profile journals. As a regular reviewer for *Nature*, I find that submissions often contain some ‘bold claim’ about the extraordinary novelty of the results presented. These bold claims are designed to get the paper into the review process. In my experience, reviewers often find that the work in the manuscript is good science and quite interesting, but that the bold claim typically cannot be justified by the data that have been presented. The authors then dilute or remove the bold claim, with the result that a good, quite interesting manuscript is published, but is no better than many simultaneously being published in specialized journals. The difference is that the *Nature* submission began with hyperbole and overselling, the traces of which have vanished from the final version. There is a danger that authors are rewarded for a fundamentally dishonest approach.

Second, the review process has no power to screen for false authorship. Often, a large laboratory is likely to get one or more papers in high-impact

journals, and it is very easy for the group leader to insert the name of a favoured postdoc who ‘needs’ a publication for some career goal into the list of authors, notwithstanding the minor contribution that the individual made to the work. The young scientist’s CV is enhanced, and the rewards follow, not just for the young scientist, but also sometimes for the group leader.

John Brookfield

Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK

Nature’s authorship policy is that authors are strongly encouraged to include a statement in the Acknowledgements to specify the actual contribution of each co-author. See www.nature.com/nature/submit/policies/index.html#2 — Editor, Correspondence

Impact factors aren’t top journals’ sole attraction

Sir — Peter A. Lawrence (*Nature* **422**, 259–261; 2003) points out in his Commentary that a substantial amount of politics has become commonplace in the scientific-publication enterprise.

In the numerous fields that are a bit less brutal than the highly competitive biomedical sciences, there may be another compelling motivation for authors to seek publication in the highest-ranked journals. The top journals are the most likely to consult top referees successfully, thus providing authors with the most meaningful feedback. The problem of fierce competition and conflicts of interest, although undeniably present, is probably less severe in fields such as the geosciences, and the benefit of receiving timely, top-notch reviews — rather than off-the-point comments from perhaps less qualified referees — should not be underestimated.

Torbjörn E. Törnqvist

Department of Earth and Environmental Sciences, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607-7059, USA

Disruption to science in developing countries

Sir — I fully agree with the concern expressed by Peter A. Lawrence in his Commentary (*Nature* **422**, 259–261; 2003) about how different strategies and manoeuvres adopted by scientists desperate to publish in a few top journals can be disruptive to the quality of research. But in addition, the policy of considering

“the journal to be more important than the scientific message” is having an even more devastating effect on science in developing countries.

I am speaking not only of my experience in Brazil, but also of other countries of similar social and economic standing, where money — and money for science — is much scarcer than in developed countries. Why is it worse here than there?

First, of course, there is much less money to be invested in science, and thus its misuse is proportionally more detrimental. Second, the ‘accountability culture’ has been imported and widely adopted, with neither assessment of its validity nor critical analysis of its consequences. It is troublesome that such an essential issue has been taken for granted by the scientific community, which by definition is supposed to accept facts and procedures only when solid data leave no space for doubt.

Finally, a numerical assessment of scientific merit minimizes the number of important variables, and consequently reduces the possibilities of defining the priorities and scientific strategies that best fit local demands.

I am fully aware of the international character of science and of the universality of the criteria for judging its quality. However, this global view does not mean that each country should forsake its individuality and its capacity to define its own criteria and methods for assessing merit.

Marcello A. Barcinski

Department of Parasitology, University of São Paulo, Av. Lineu Prestes 1374, São Paulo 05508-900, Brasil

Separate achievements of the Humboldt brothers

Sir — In contrast to the assertion made in your Editorial “Berlin’s university crisis” (*Nature* **423**, 101; 2003), the famous principle of the “unity of teaching and research” that forms the conceptual basis of modern academic education was not developed by the natural scientist Alexander von Humboldt (1769–1859); neither was he the “founder of the modern German higher-education system”.

These achievements were those of his brother, the politician and linguist Wilhelm von Humboldt (1767–1835), who in 1810 founded the university of Berlin that carries his name today.

W. S. Peters

Institut für Allgemeine Botanik, Justus-von-Liebig-Universität, Senckenbergstrasse 17-21, D-35390 Gießen, Germany

Beneath the great divide

What was life on Earth like before the Cambrian explosion?

Life on a Young Planet: The First Three Billion Years of Evolution on Earth

by Andrew H. Knoll

Princeton University Press: 2003. 304 pp.

\$29.95, £19.95

Stefan Bengtson

There is a 'great divide' in the fossil record, known as the Cambrian explosion. Above it is a plethora of creatures large and small; below it are mostly microbes and slime. Palaeontologists spend 99% of their efforts working on fossils above it — but what do we know about what's below? Is the great divide what we think it is?

The prominent Harvard palaeontologist-geobiologist Andrew Knoll is an ideal guide through this early phase of life's history on the Earth. ('Early' is a relative term here, for when his narrative ends, the 'young' planet is four billion years old and has 'only' half a billion years to go before the *Beagle* sets sail from Devonport.)

Knoll starts and ends (barring a final trip into astrobiological space and a pleasing philosophical epilogue) with the Cambrian period. The reason is pedagogic: first introduce the reader to a world that is at least vaguely familiar, then plunge into deep time to find the beginnings of things, gradually returning to the more familiar territories. Just as the subject is broad and varied, so too are the demands on the reader: the text moves between impressionistic images of fieldwork and technical accounts and explanations. The reader has to be prepared to change gear, and if you don't know what a Rube Goldberg machine is, you could always try the Internet.

The dearth of data in Precambrian palaeobiology has inspired lateral thinking and multidisciplinary approaches — but it has also made the field vulnerable to speculation and wishful thinking. Even though the number and quality of Precambrian fossil sites have grown significantly during the past half-century, the available information is still scanty and scattered compared with that above the great divide. How representative is it?

Science is the art of not fooling yourself (as Richard Feynman put it), but the lure to present a coherent story by connecting dots is strong. None of us is immune to this, and one of the strengths of Knoll's book is that it presents science as the open-ended endeavour that it is. In one of the final chapters he writes: "the absence of a definitive punch line ... is why I get up in the morning." Ah, yes, what could be more depressing than



Hot springs eternal: cyanobacteria exist in conditions that may not have changed for billions of years.

knowing that all of the problems are solved! Knoll sees the beauty of science as a never-ending story, and transmits his vision well to the reader.

So, despite the remarkable progress of the past 50 years, some of the most crucial questions in Precambrian palaeobiology are still unanswered. When, where and how did life begin? The evidence for a very early establishment of life on Earth is currently under heavy fire; whether life started hot or cold, in soup or stone, on Earth or beyond, is undecided. Two other long-standing enigmas are why it took a billion years or more for eukaryotes to start making their mark after the prokaryotes appeared, and why another billion or two passed before multicellular life finally erupted in the Cambrian explosion. Did evolution really need so long to build complex creatures, or was something keeping the lid on those cans of worms during eons of enforced tranquility?

The Knoll hypotheses regarding evolution's dallying and rallying are 'oxygen' and 'permissive ecology'. The ideas are not new, but Knoll has been instrumental in developing them. The first great oxygenation event in the Earth's atmosphere identified by geochemists took place about 2.2 billion years ago, with a second just less than 1 billion years ago. Knoll believes that the latter event fuelled the Cambrian explosion.

A prerequisite is permissive ecology — an

initial lack of fierce competition, allowing for rampant diversification before developmental pathways and tightening ecological interactions restricted the possibilities. Although Knoll believes that preceding extinctions thus paved the way for the Cambrian explosion, he is lukewarm about the Snowball Earth hypothesis (the idea that the whole Earth froze over), which would otherwise be a pretty good extinguisher in this scenario. Slushball, yes; snowball, no.

What did the earlier oxygenation event result in? Knoll argues that this, too, led to an expansion of the ecology, where oxygen-dependent organisms, including eukaryotes, caught the whiff of a grander future. Eukaryotic evolution after this event seems remarkably lethargic, however. Maybe the reason for this is what Knoll calls the 'Canfield' ocean — the global Black Sea that geochemist Don Canfield has proposed to be the result of moderately low ambient oxygen levels following on from the earlier boost. If so, life for the early eukaryotes may have been a harsh struggle against limiting conditions for more than a billion years. And if organisms and rocks are scarce, we need to search long and hard to acquire a data set that we can trust to give us a complete picture of the Precambrian biotas.

The mark of a healthy research field is that there is never a good time to write a book about it. Like other attempts to spell out the

history of early life on Earth, *Life on a Young Planet* is a time document, but it expresses better than most the bumptious vitality and sheer fun of open-minded research. ■

Stefan Bengtson is in the Department of Palaeozoology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden.

More on the Cambrian explosion In the Blink of an Eye

by Andrew Parker
Perseus, \$24.95, Can\$ 38.95; *Free Press*, £18.99

The journey to Enlightenment

The Man Who Flattened the Earth: Maupertuis and the Sciences in the Enlightenment

by Mary Terrall
University of Chicago Press: 2002. 408 pp.
\$39, £27.50

Steven Shapin

Being Pierre-Louis Moreau de Maupertuis was hard work. Some of that work was deciding what kind of man of science he wanted to be. Starting out in mathematics and mechanics, he eventually made notable contributions to the study of heredity, and ended his career swept up in a series of metaphysical disputes. Some of the hard work was physical: his most notable scientific research in geodesy called for courage and a willingness to suffer for science. In 1736, he led an expedition to Lapland to establish the precise measure of a degree of latitude. Heavy astronomical instruments had to be lugged up mountains at the northern end of the Gulf of Bothnia, and measuring rods were laid end-to-end for miles over frozen rivers. In winter

the cold was so cruel that the intrepid group's cups of eau-de-vie froze to their lips, and when summer came, swarms of mosquitoes drove them to the brink of madness.

The expedition to Lapland was a display of manly fortitude as much as an exercise in state-funded scientific rationality. When Maupertuis returned to Paris, he ensured that the world properly appreciated his achievements by publishing a vivid account, imaginatively combining the genres of travelogue, 'boy's own' adventure story and popular science.

Maupertuis had survived the rigours of Lapland, but now he had to make credible the precision knowledge brought back from the Arctic. What was at issue in the measure of equatorial and Arctic degrees of latitude was the bitterly contested question of whether the shape of the Earth was oblate (flattened at the poles), as Maupertuis maintained, or prolate (elongated at the poles), as was claimed by his local enemies in the Paris Academy of Sciences, the Cassini dynasty of astronomers. The problem of the Earth's exact shape was pertinent to practical cartography and navigation, but it was also connected in complicated ways to preferences for newtonian or cartesian theories of gravitation in particular, and philosophies of nature in general.

For Enlightenment men of science working in absolutist political settings, the question of the credibility of scientific knowledge was always linked to the security of their scientific careers. Science played to the sovereign and his courtiers, and without their approval there were limited resources to do research, few institutional stipends to sustain them, and only weak or marginalized cultural allies to support scientific positions. Part of Maupertuis' genius lay in the deft way that he kept audiences of scientific experts and polite society both in play. As Mary

New in paperback

Dr Tatiana's Sex Advice to All Creation: The Definitive Guide to the Evolutionary Biology of Sex

by Olivia Judson
Owl/Metropolitan/Henry Holt, \$14;
Vintage, £7.99

"Judson's text is both wonderfully entertaining and authoritative. All in all this is a stimulating feast of extraordinary sexual practices." Tim Birkhead, *Nature* **418**, 483 (2002).

Shoemaker by Levy: The Man Who Made an Impact

by David H. Levy
Princeton University Press, \$16.95, £11.95
"Levy writes well, and his pacey style keeps his personal story bubbling along superbly. The book is everything a 'good read' should be." David W. Hughes, *Nature* **409**, 133-135 (2001).

The Secret Life of Dust: From the Cosmos to the Kitchen Counter, The Big Consequences of Little Things

by Hannah Holmes
Wiley, US\$14.95, Can\$23.50
Shortlisted for the Aventis Prize for Science Books 2002.

Terrall writes, after many years, Maupertuis refined "an identity as a public figure equally at home in academy and salon".

When he returned from Lapland, Maupertuis was celebrated by both the court and the salons, falling into the arms of a series of married aristocratic literary ladies, whom he serenaded on his guitar, and securing a pension from the crown for his services to the state. He had his portrait painted dressed in exotic Lapp costume, with his left hand flattening the North Pole of a globe, just so no one could possibly miss the point. He never forgot the importance of amusing as well as instructing, and his literary merits were certified in 1743 by election to the Académie française, thus becoming an 'immortal'; and one of the very few French men of science honoured by the élite academies of both of the 'two cultures'.

Having achieved all he could in Paris, Maupertuis was now tempted by Louis XV's absolutist rival, Frederick the Great, who wished his court in Berlin to be furnished and burnished by the best Enlightenment literati that money could buy. Frederick wanted Maupertuis to take charge of the Berlin Academy of Sciences, but the Prussian king's first approaches were rebuffed. Maupertuis eventually succumbed in 1745, seduced by the opportunity for dinner-table intimacy with a monarch, by the offer of total authority over the academy's affairs, by piles of cash, and possibly by Frederick's arrangement of a posh German wife for



Snow easy life: Maupertuis endured the Arctic winter in his quest to determine the shape of the Earth.

Science in culture

Deacon's nucleic acid

A DNA anniversary project took shape in an unpredictable way.

Martin Kemp

Major artistic projects, like scientific research programmes, can develop in unanticipated directions, although in this era of predefined outcomes for grant-funded projects in academia, the truly unpredictable is increasingly factored out. Where there is no contract with a funder, however, the constraints are absent — but so, generally, are the necessary resources.

These remarks preface the story behind the genesis of an important piece of sculpture recently completed by British artist Richard Deacon in response to a scheme to celebrate the 50th anniversary of Watson and Crick's seminal discovery of the structure of DNA. In 2001, Philip Campbell, the editor of *Nature*, approached me and my colleague Marina Wallace with the idea of stimulating a substantial piece of sculpture to commemorate the anniversary. None of us wanted a straightforward representation of the double helix, which has entered the world of visual cliché. Reviewing possible artists, we settled on Deacon, who had an impressive record of generating large-scale constructions, especially in wood, that look like 'real' things from the organic world but are in fact entirely invented.

Deacon took up the challenge enthusiastically, undertaking background research. The Gulbenkian Trust generously offered set-up money, and a full budget proposal was developed. Surprisingly, no major funder came forward.

But Deacon was undeterred. Carried forward on the wave of his creative impetus, he assumed the risk himself — and broke free of constraints. The developing work began to take on a fascinating life of its own in a process of visual and physical dialogue with its creator.

Deacon commissioned the making of a series of strongly bent and fiercely twisted oak components according to a set repertoire of barley-sugar twists, and 'fast' and 'slow' curves. His demands literally stretched the wood to its limits. The torsions were not the result of predetermined pressures in a machine but arose from progressive forces instinctively applied to the steamed timbers by two human operatives, whose eyes, ears and hands provided the most sensitive registers of when breaking point was near.

The components were at the centre of an inter-



DNA with a twist: Richard Deacon's *Out of Order* has evolved its own 'molecular' structure.

play between Deacon's visualization, the wood's inherent structural possibilities, and what seemed to be the self-organizational needs of the growing sculpture. Originally, the parts were arranged into four separate but related wholes, like the subunits of a large molecule. Then, suddenly, they came together, bolted and bonded, with a meandering and twisted spine from which sprays of ribbons launch themselves into space at skewed angles, seeking connections. The sculpture seems both closed and open, depending on the viewpoint.

The result, without any defined intention, is uncannily evocative of a ribbon model of a large protein molecule or nucleic acid as generated by a graphics programme such as MolScript. It is at once precariously dynamic and structurally poised. A piece of sculpture as large as this (700 × 570 × 190 cm) writhing in a confined space could well feel threatening, but those who witnessed its unveiling at the Lisson Gallery on 1 May could be seen smiling at its joyous vitality.

Deacon's magnificent sculpture is no longer about DNA. Its title is metamorphosing from the genetically orientated *Out of Africa* to the more

abstract *Out of Order* (suggesting, among other things, the emergence of unpredictable complexity from a limited set of basic parts). But its rootedness in a conceptual process akin to that which allowed Watson and Crick to see the structure of DNA remains central to its formation.

Where now for the sculpture? In the short term it is to cast its spell in Tampere in Finland and Vitoria in Spain, where it will no doubt prove itself to be a show-stopper. Where it will find a permanent home is still to be determined.

Martin Kemp is professor of the history of art at Oxford University and co-director of Wallace Kemp/Artakt. His book *Visualizations: The Nature Book of Art and Science* is now available in a German edition with the title *Bilderwissen: Die Anschaulichkeit Naturwissenschaftlicher Phänomene* (Dumont Literatur und Kunst Verlag, 2003).

Out of Order will tour European venues next year in a major exhibition of Richard Deacon's work, starting at the Sara Hildén Art Museum in Tampere.

the now-47-year-old French bachelor savant.

Exile to the Prussian cultural hinterlands turned out to have limited charms for Maupertuis. He never mastered the language; he was irritated by German philosophers' preference for Christian Wolff's version of leibnizian metaphysics; and he got caught up in a nasty, and ultimately petty, polemical exchange with Voltaire, Frederick's other prize French catch. His health already suffer-

ing from the Prussian climate, Maupertuis faced personal disaster when Frederick and Louis, the two absolutist sovereigns to whom he owed allegiance, went to war with each other in 1756. Maupertuis died three years later, fittingly in Basel, about halfway between Paris and Berlin.

Maupertuis' life exposes some of the tensions that existed between the cultural powers of princely patronage and Enlighten-

ment ideals of intellectual republicanism. This well-crafted biography is one of the better studies of the problems and opportunities of the eighteenth-century scientific career under the conditions of absolutism. ■ Steven Shapin is in the Department of Sociology, University of California, San Diego, La Jolla, California 92093-0533, USA. His books include *A Social History of Truth: Civility and Science in Seventeenth-Century England*.

In search of clarity

Gautam R. Desiraju

Concepts that are ancient and pervasive do not lend themselves easily to precise definition. A layperson might think of 'crystal' as referring to table glass of high quality. But to a scientist, 'crystal' denotes solid matter in a more or less ordered form, whereas glass is considerably less ordered than a truly crystalline substance. Yet the Spanish, for example, have only one word (*crystal*) for both 'glass' and 'crystal'.

The dictionary definition of crystal is clinical, referring to "forms assumed by substances with a definite internal structure and external shape of symmetrically arranged plane surfaces". But the idea that crystals are cold and static has persisted since the time of ancient Greece — the term *crystallos* ('ice' or 'quartz') evolved from *cryos* ('cold') and *halas* ('salt'). As recently as the mid-1900s, the Nobel prizewinner Leopold Ruzicka dismissed crystals as "chemical cemeteries". Crystals were dead and so no chemistry could take place within them.

Physicists have taken a phenomenological approach. According to the International Union of Crystallography, a crystal is any solid that gives a discrete X-ray diffraction diagram. A crystal is thus defined not by what it is but by how it appears. But when an entity is defined by an observed property, its very existence depends on the method of observation. Crystals were first observed by eye, then by microscope, then by X-rays. What will be next?

The result of an X-ray structure analysis is a crystallographic unit cell and its constituent atoms. These supposedly identical building blocks are repeated along three principal axes (an arrangement known as periodicity) and crystallographers speak of

crystals as entities with long-range order. But this unit cell is an averaged measurement. No two unit cells are exactly the same and no unit cell looks exactly as it did a moment ago. If we are able to examine the contents of a single unit cell at a fixed point in time, will we change our definition of a crystal? The present definition was prompted by the discovery of quasicrystals. These have long-range order but lack the structural periodicity of classical crystals. Will there be other new forms of long-range order? Do sharp spots and a discrete diffraction pattern imply order? Already we know of giant biological crystals, such as ribosomes, that produce typical diffraction patterns despite their lack of strict periodicity. There is also evidence that if very intense X-rays are used, diffraction could be observed for non-crystalline materials. Indeed, crystallography today looks through a small window at what is potentially an expansive landscape.

Chemistry provides a more pragmatic definition of a crystal — bridging the divide between the austerity and accuracy of physics and the exuberance and excitement of biology. To a chemist, crystallization brings atoms, ions or molecules together to form a condensed phase that is characterized by some degree of order. A crystal is a manifestation of mutual recognition, a storage device for structural information and a victory of enthalpy over entropy. From the fluid state, in which molecules move randomly and ceaselessly, to the ordered cloister that is the crystal is a long journey, and one of the most remarkable reactions in all of chemistry. In this supramolecular reaction, it is weak intermolecular interactions, rather than strong covalent bonds, that are made and broken.

The emergence of a crystal from a fluid may seem magical, but closer inspection brings us nearer to discovering the energetic secrets of the process. There could be a gradual withdrawal of entropy and an ascendancy of enthalpy — a shift in the 'enthalpy-entropy balance' — just before the formation of the crystal nucleus, the free-energy maximum between the fluid and the crystal. Even as molecules begin to assemble through the mediation of weak intermolecular forces, giving rise to hydrogen bonds, electrostatic and van der Waals interactions, it is possible that the solute-solvent assemblage retains some 'wobble', so that the free energy of nucleation is reduced both by a favourable enthalpy of nucleation (formation of intermolecular bonds) and a less unfavourable entropy of nucleation (mobility of the solute-solvent cluster). In this way, order enters the putative crystal.

Crystal

Scientists differ from the public, and even among themselves, in their ideas of what constitutes a crystal. As these varying approaches testify, the definition of this form of matter is not as clear as its name suggests.

As in other chemical reactions, the products of the crystallization reaction are messy. So, rather than producing geometrically perfect constructs, the process yields crystals with defects, dislocations and incommensurate subdomains. Biological crystals contain so much water that different portions can have different flexibilities and mobilities. Metastable crystals and nanocrystals are kinetically trapped, whereas polymorphic crystals can be elusive. Even three-dimensional order is unnecessary — membrane protein arrays, for example, are ordered in only two dimensions, and fibres in only one.

But what of order, which seems to be a characteristic of all crystals? For most crystallographers it is long-range order that sets crystals apart from glasses and amorphous solids. But condensed-matter physicists have recognized that there is a structural continuum between ideal crystals and amorphous solids, and have quantified this concept with the order parameter. A variety of ordering can, in principle, be sampled as one moves along the reaction coordinate of crystallization. There is evidence that an amorphous phase might well be an intermediate stage en route to the formation of a more perfectly ordered crystal. Perhaps the Spaniards were right all along.

It seems that the issue is not how one defines crystal but how one should define order. Terms such as 'crystalline', 'amorphous', 'periodic', 'quasiperiodic' and 'aperiodic' only reveal our helplessness as we stumble around in small regions of a vast structural panorama. Scientists have long assumed that long-range order in a periodic crystal represents the lowest possible free energy for condensed phases, but no one has actually proved this. As chemistry moves towards systems of greater complexity and diversity, the meaning and scope of the term 'crystal' can only evolve and expand. ■

Gautam R. Desiraju is at the School of Chemistry, University of Hyderabad, Hyderabad, 500 046, India.

FURTHER READING

Sayre, D. *Struct. Chem.* **13**, 82–96 (2002).
Lehn, J.-M. *Supramolecular Chemistry: Concepts and Principles* (VCH, Weinheim, 1995).
Desiraju, G. R. *Crystal Engineering: The Design of Organic Solids* (Elsevier, Amsterdam, 1989).



Beautiful demonstration: but not all crystals are as highly ordered as the Koh-i-Noor diamond.

Oestrogen receptor hijacked

Jan J. Brosens and Malcolm G. Parker

Widespread pollution of the environment by dioxins poses a risk to human health. The mechanism used by these chemicals to alter the body's responses to oestrogens is now being unveiled.

The extent to which environmental pollutants cause human disease has preoccupied and divided the scientific community for decades¹. Dioxins in particular are often singled out for discussion, having been blamed for disorders ranging from cancer to immune deficiencies, and from fetal abnormalities to the perceived decline in male and female fertility. The idea that these chemicals affect fertility is largely based on animal studies, which show that dioxins are potent hormone disruptors, and can, in different circumstances, either promote or antagonize the effects of oestrogens². On page 545 of this issue, Ohtake and colleagues³ look at how dioxins affect cellular responses to these hormones.

Oestrogens are steroid hormones that help to regulate the growth, differentiation and function of reproductive tissues, as well as those of other organs such as the bones, brain and cardiovascular system⁴. Key to these diverse effects is the activation of the oestrogen receptors ER α and ER β — members of a superfamily of receptors found in the cell nucleus — by their hormonal ligands, which include 17 β -oestradiol (Fig. 1a). The ligand-activated receptors bind to the control regions (promoters) of specific genes, where they recruit co-activators or co-repressors and the machinery that brings about gene expression. The receptors thereby control the transcription of target genes and elicit a cascade of cellular events.

Mammalian cells also contain a variety of receptors that can bind and respond to toxic compounds present in the environment. These receptors include other members of the nuclear receptor family, such as PXR and CAR, as well as the unrelated aryl hydrocarbon receptor (AhR)^{5,6}. Dioxins and related pollutants selectively bind AhR, and recruit a chaperone protein, the aryl hydrocarbon nuclear translocator (Arnt; Fig. 1b). The resulting transcriptionally active complex induces the expression of a battery of genes, including those that encode detoxification enzymes.

Now, through a series of experiments in cultured breast and uterine cancer cells, Ohtake *et al.*³ show that the binding of dioxins to AhR can also cause this receptor to associate directly with oestrogen receptors (Fig. 1c). In this way, dioxins can activate oestrogen receptors in the absence of 17 β -oestradiol, and so switch on the expression of

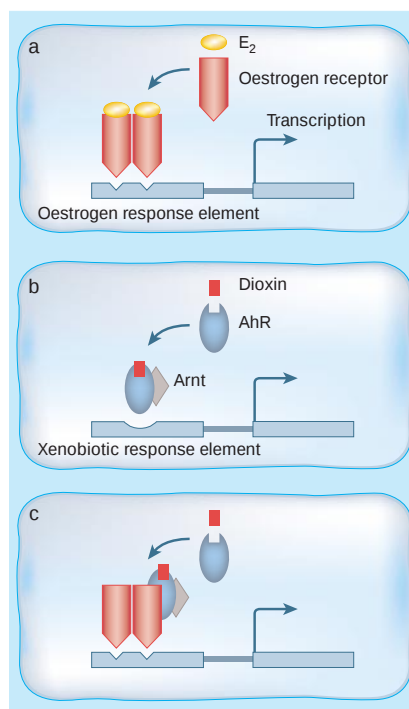


Figure 1 Dioxins take oestrogen receptors for a ride. a, Oestrogens such as 17 β -oestradiol (E₂) regulate cellular processes by binding to oestrogen receptors in the nucleus. The activated receptors dimerize and bind to specific oestrogen response elements in the promoter regions of target genes, thereby triggering gene transcription. b, The aryl hydrocarbon receptor (AhR), when activated by dioxins, forms a complex with the aryl hydrocarbon nuclear translocator (Arnt) that can also trigger transcription. The complex regulates gene transcription by binding to specific elements in DNA, termed xenobiotic response elements. c, Ohtake and colleagues³ have discovered that dioxins can mimic the effects of oestrogens through a mechanism that involves the activation of oestrogen receptors by a transcriptionally active AhR–Arnt complex.

oestrogen-responsive genes. In contrast, when 17 β -oestradiol is present, activation of AhR by dioxins impairs the expression of oestrogen-responsive genes. This latter effect of dioxins — and other selective AhR modulators — has also been reported previously⁷, and may result from an enhanced degradation of liganded oestrogen receptor when it is bound to AhR.

Ohtake *et al.*³ extended their *in vitro* observations by injecting dioxins into mice that had had their ovaries removed — and which therefore lacked circulating 17 β -oestradiol. They found that the dioxins induced the expression of several of oestrogen's target genes, enhanced the proliferation of cells in the glandular compartment of the uterus, and increased uterine weight. Such responses were not observed in mice that had been genetically manipulated to lack either the AhR or the ER α gene, emphasizing the importance of both receptors in mediating the 'oestrogenic' effects of dioxins.

What are the implications of Ohtake and colleagues' observations? From a molecular viewpoint, this study reveals another way in which a member of the steroid-receptor family can be activated in the absence of its ligand. Other such mechanisms have been described. For instance, oestrogen receptors have been shown to be activated by other hormone-bound steroid receptors^{8,9}. Moreover, ER α can be activated by certain growth-factor proteins in the absence of oestrogen, both *in vitro* and *in vivo*¹⁰. These findings are supported by a study of mice engineered to express a luciferase gene — the product of which can be detected by light emission — that was under the control of activated oestrogen receptors⁴. The expression of luciferase did not depend solely on the levels of circulating ovarian oestrogens, especially in non-reproductive tissues.

Why are there so many diverse mechanisms for activating the oestrogen receptor? An evolutionary survey¹¹ suggests that this receptor is the likely ancestor of all other steroid receptors in vertebrates. It is conceivable that the activity of the ancestral oestrogen receptor was controlled initially by molecular signalling cascades rather than by oestrogens, given that the nuclear receptor family is also known to be regulated by phosphorylation, a common molecular signal. The acquisition of oestrogen-dependency may have been an evolutionary process that allowed reproductive events in the body to become synchronized.

Similarly, AhR arose more than 450 million years ago, well before the environment became universally contaminated by man-made compounds⁶. Several lines of evidence support the notion that AhR may have functions beyond its role as a chemical sensor and defence molecule. For instance, it has a

promiscuous ligand-binding site and can potentially be activated by many naturally occurring dietary and endogenous compounds⁶. Furthermore, the characteristics of AhR-deficient mice suggest a role for this receptor in the immune and vascular systems, the liver, ovaries and other organs^{5,12,13}. Together, these observations raise the possibility that the activation of AhR by natural ligands, and its interaction with oestrogen receptors³, may be important during development and perhaps in adult life.

Clearly, Ohtake and colleagues' findings also provide a mechanism whereby exposure to chemicals can disrupt normal oestrogen-induced signalling events. Dioxins, the by-products of many industrial and combustion processes, have been detected in the blood of virtually every person tested¹⁴. They are present in the fat of breast milk, and can cross the placenta from mother to fetus. Toxicology studies indicate that exposure to low levels of toxins during critical developmental periods can have permanent adverse effects on health. To make matters worse, the current burden of dioxins and other chemicals in certain human populations has been reported to be near the range at which adverse effects occur in laboratory animals^{2,14}.

Given these facts, why is it so difficult to establish an unequivocal causal link between dioxins and human disease in the general population? Part of the problem is that many

of the toxic effects attributed to dioxins — such as impaired immunity, fertility and cognition — are relatively subtle, multifactorial, difficult to diagnose and quantify, and often take years to appear. Resolving these issues will depend on a combination of improved diagnostic tests, large epidemiological studies, and the development of genetically manipulated animal models in which the role of crosstalk between the AhR and oestrogen-receptor signalling cascades can be distinguished from the role of their individual pathways. ■

Jan J. Brosens and Malcolm G. Parker are at the Institute for Reproductive and Developmental Biology, Imperial College London, Hammersmith Campus, London W12 0NN, UK.
e-mail: m.parker@imperial.ac.uk

1. Axelrod, D., Davis, D. L., Hajek, R. A. & Jones, L. A. *Environ. Health Perspect.* **109**, A246–A249 (2001).
2. Rier, S. & Foster, W. G. *Toxicol. Sci.* **70**, 161–170 (2002).
3. Ohtake, F. *et al.* *Nature* **423**, 545–550 (2003).
4. Ciana, P. *et al.* *Nature Med.* **9**, 82–86 (2003).
5. Hahn, M. E. *Chem. Biol. Interact.* **141**, 131–160 (2002).
6. Denison, M. S. & Nagay, S. R. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334 (2003).
7. Wormke, M. *et al.* *Mol. Cell. Biol.* **23**, 1843–1855 (2003).
8. Migliaccio, A. *et al.* *EMBO J.* **17**, 2008–2018 (1998).
9. Koustini, S. *et al.* *Cell* **104**, 719–730 (2001).
10. Klotz, D. M. *et al.* *J. Biol. Chem.* **277**, 8531–8537 (2002).
11. Thornton, J. W. *Proc. Natl Acad. Sci. USA* **98**, 5671–5676 (2001).
12. Fernandez-Salguero, P. *et al.* *Science* **268**, 722–726 (1995).
13. Lahvis, G. P. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 10442–10447 (2000).
14. Thornton, J. W., McCally, M. & Houlihan, J. *Public Health Rep.* **117**, 315–323 (2002).

to produce HD; if instead the O–H bond was excited, H₂ was produced.

But similar experiments in reaction control are rarely successful in condensed phases (liquids, glasses and solids) because a specific vibrational excitation becomes redistributed rapidly over other modes within the molecule or among its neighbours. To decipher the effect that the local environment exerts on an individual molecule, investigators turned their attention to experiments on single molecules. For example, Xie and Dunn³ recorded the fluorescence spectrum of a single molecule on a glass surface, and were able to discern fluctuations in the molecule's local environment that could not be seen in the data recorded for an analogous bulk sample.

Another line of approach in single-molecule spectroscopy involves the scanning tunnelling microscope (STM), invented in the 1980s by Binnig and Rohrer. An immense collection of solid-surface images has since been produced by STMs, with sufficient resolution to reveal the atomic and molecular adsorbates that might decorate a surface. Then in 1998, Stipe *et al.*⁴ made the remarkable discovery that the current flowing from the tip of an STM can selectively excite different modes of vibration in a lone molecule that is bonded onto a surface.

Other investigations have shown that an STM tip can be manipulated to translate or rotate⁵, or fragment⁶, an atom or molecule bonded to a surface — or even to eject it from the surface⁷ (desorption). An STM tip can be used to force two molecules to dissociate, and the resulting fragments can be rearranged and fused together to form a new product molecule⁸. In each of these cases, however, the STM tip has served either as an atomic-scale poker to push a molecule physically across the surface or as a localized heater (through electron bombardment) for inducing non-selective thermal excitations.

Pascual *et al.*¹ now demonstrate another way in which the STM can control the behaviour of a molecule on a surface. They use an STM tip to selectively excite vibrational modes in an ammonia molecule (NH₃) as a way to sever the chemical bond between NH₃ and a copper surface, or alternatively to induce the molecule to move laterally across the surface (Fig. 1). The reaction outcome is controlled simply by selecting the tip voltage and thereby setting the energy of the tunnelling electrons to a value that will resonantly excite the preferred vibrational mode. Of the two modes investigated, the first is a stretching, or 'breathing', mode in which all three N–H bonds are symmetrically stretched and compressed, the second a bending mode that resembles an umbrella inverting on a windy day. Most of the excited molecules relax quickly by transferring their energy to the copper lattice, but some convert their vibrational motion into translational motion. Specifically, excitation of the

Chemistry

Tips for moving single molecules

Dennis C. Jacobs

Scanning tunnelling microscopes provide a unique perspective on chemistry at the level of single molecules. Now there is a new way of using the tip of such a microscope to manipulate a single molecule.

In nature, reactants can be transformed into products through a vast array of mechanisms. In synthetic chemistry, progress hinges on finding a way to maximize the yield of the desired target molecule while minimizing the generation of unwanted by-products. In the same way that a golfer anticipates the terrain of the green when putting, chemists aim to control the chemical dynamics of a reaction, guiding the participating molecules along a specific path. On page 525 of this issue, Pascual *et al.*¹ demonstrate how a scanning tunnelling microscope can be used to control the excitations of a single molecule and achieve a desired reaction path.

Most chemical reactions require that some of the existing bonds within the reactants are weakened or broken before new bonds can form. This means that the reactants must receive a critical amount of energy from their surroundings to begin

their metamorphosis. Raising the temperature of the sample is a relatively inefficient way of doing this, because fluctuations due to thermal energy are typically ten to a hundred times smaller than the activation energy for a reaction. Furthermore, under thermal conditions, the energy is distributed over many types of molecular motion — translational, rotational or vibrational, for example — whereas usually only one specific type of motion (such as a vibrational stretch) is associated with crossing over the reaction barrier.

So a more controlled approach is needed. In 1991, Bronikowski *et al.*² showed that stretching vibrations could be excited between the atoms in molecules of deuterated water vapour (HOD instead of H₂O) by laser radiation. If the stretch mode of the O–D bond was excited, they found that an approaching hydrogen atom preferentially abstracted the deuterium atom from HOD

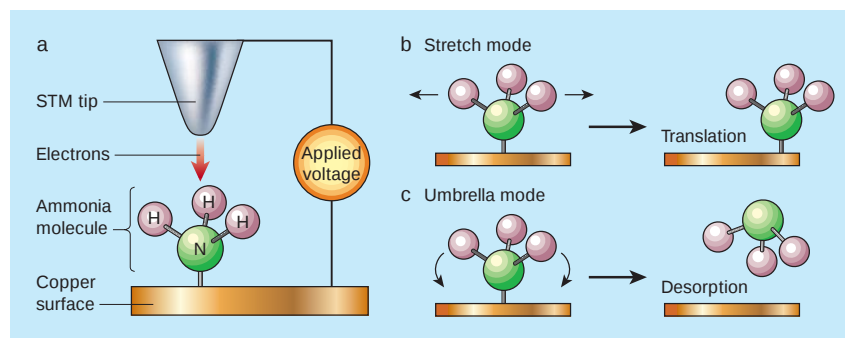


Figure 1 Moving molecules. Energetic electrons emerging (a) from the tip of a scanning tunnelling microscope (STM) can selectively excite a specific vibrational mode within a molecule. For an ammonia molecule (NH₃) adsorbed on a copper surface, electrons with an energy of 420 millielectronvolts excite a 'stretch' mode (b), causing the molecule to be translated along the surface. If the electron energy is 320 millielectronvolts, the 'umbrella' mode (c) is excited instead, and this flexing of the hydrogen–nitrogen bonds — similar to an umbrella turning inside out — results in desorption of the molecule from the copper surface.

umbrella mode tends to desorb the molecule, intact, from the surface; the symmetric breathing mode preferentially induces lateral translation of the molecule across the surface.

This pioneering experiment not only reveals how competing reaction mechanisms can be selectively activated within a single molecule, but also demonstrates a new approach for identifying reaction pathways in complex environments. Studies such as this expand the range of synthetic tools available for fabricating the next generation of molecular-scale nanostructures. ■

Dennis C. Jacobs is in the Department of Chemistry

and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA.

e-mail: jacobs.2@nd.edu

1. Pascual, J. I., Lorente, N., Song, Z., Conrad, H. & Rust, H.-P. *Nature* **423**, 525–528 (2003).
2. Bronikowski, M. J., Simpson, W. R., Girard, B. & Zare, R. N. *J. Chem. Phys.* **95**, 8647–8648 (1991).
3. Xie, X. S. & Dunn, R. C. *Science* **265**, 361–364 (1994).
4. Stipe, B. C., Rezaei, M. A. & Ho, W. *Science* **280**, 1732–1735 (1998).
5. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524–526 (1990).
6. Stipe, B. C. *et al.* *Phys. Rev. Lett.* **78**, 4410–4413 (1997).
7. Shen, T.-C. *et al.* *Science* **268**, 1590–1592 (1995).
8. Hla, S.-W., Bartels, L., Meyer, G. & Rieder, K.-H. *Phys. Rev. Lett.* **85**, 2777–2780 (2000).

Evolutionary biology

Fractional phylogenies

Thomas D. Kocher

Speciation has been unusually fast among the cichlid fishes of Lake Victoria. An unexpectedly distant ancestor, which perhaps already had a predisposition for rapid speciation, may have seeded this 'species flock'.

More than half of all living species of vertebrates are 'bony' fishes. Over the past 250 million years they have evolved into more than 25,000 species that span habitats from the highest mountains to the deepest oceans. Nowhere has their evolutionary radiation been more rapid than in the great lakes of East Africa. Several hundred species of cichlid fishes swim in the waters of Lake Victoria, and it has been assumed that this 'species flock' arose within the lake from one or a few common ancestors since the lake last dried up around 15,000 years ago¹. In a paper in *Science*, Verheyen and colleagues² cast doubt on this assumption, and instead suggest that Lake Victoria was colonized several times, beginning more than 100,000 years ago, by fish from Lake Kivu, which lies 300 km to the west.

The very recent origin of the Lake

Victoria flock poses two challenges for those wishing to reconstruct the historical relationships among species. First, there has been little time for mutation to alter the DNA sequence of each species. This means that there are precious few sequence characters on which to apply phylogenetic analysis. Second, there has not been enough time for new variant genes to become fixed between instances of speciation³, a problem known as 'incomplete lineage sorting'. This means that although phylogenetic trees derived from DNA sequences accurately represent the history of genes, they do not necessarily reflect the history of the populations in which the variants are found (Fig. 1, overleaf).

The problem of too few mutations is usually addressed by focusing on rapidly evolving gene sequences. DNA in the cell's mitochondrion has been a particular

favourite for such studies, because it evolves roughly ten times faster than the average gene in the nucleus of the cell. An earlier study of the Lake Victoria flock⁴ found considerably more diversity of mitochondrial DNA than could have arisen since the lake last dried out 15,000 years ago. This means that the most recent radiation in Lake Victoria was seeded by several lineages, in which genes for important morphological and behavioural traits were already polymorphic.

The bit added by Verheyen and colleagues² is data on the mitochondrial DNA sequences of cichlids native to Lake Kivu, which lies quite far to the west of Lake Victoria. They construct a 'haplotype network' in which each sequence is joined to the next in a way that explains the sequence differences with a minimum of mutational events. Then they consider the frequency and geographical distribution of these haplotypes so as to reconstruct a plausible biogeographical scenario for the evolution of the fishes. Their analysis identifies the Lake Kivu species *Haplochromis gracilior* as the closest relative of the Lake Victoria superflock, and suggests that there has been a complex pattern of faunal exchange through the region.

Neither of these studies^{2,4} overcomes the problem of incomplete lineage sorting, and so the gene trees must be interpreted with caution. An alternative is to score DNA polymorphisms at thousands of independent gene loci from the cell nucleus⁵. The average of these many gene trees should be an accurate estimate of the history of the populations in which the genes are segregating. Seehausen and colleagues⁶ have used this second approach to analyse a number of potential riverine ancestors, and have identified several species of the genus *Thoracochromis* as the closest relatives of the Lake Victoria superflock.

The nuclear and mitochondrial studies provide complementary perspectives, but it is difficult to combine their results into a coherent whole because the various data sets do not overlap for certain key species. Even if complete data were available, it is not clear that convincing statistical support would emerge for either phylogenetic tree because of the short time-frames involved. Verheyen *et al.* report a Bayesian posterior probability of 80% in support of their main conclusion — the linkage of *H. gracilior* to the Lake Victoria superflock — but such probabilities have been shown to be excessively liberal⁷. Seehausen *et al.* make the uncomfortable suggestion that the short branches in their nuclear gene tree may reflect a period of extensive hybridization among species early in the history of the flock. In that case it may never be possible to completely reconstruct the relationships among these species.

Regardless of the details, none of these analyses refutes the conclusion that most of the current species diversity has arisen within

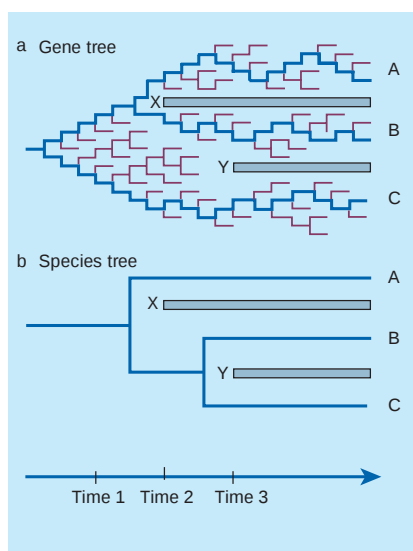


Figure 1 The difference between gene trees and species trees. a, An ancestral species contains many mitochondrial DNA variants (time 1). The evolution of a species barrier (X) separates the ancestral population into two gene pools (time 2). A second species barrier (Y) arises later, producing the full set of species (A, B, C). Phylogenetic reconstruction from the mitochondrial gene sequences suggests that species A and B are most closely related. But as shown in b, B and C are the most recent and closely related species, having shared a common gene pool until time 3.

the confines of Lake Victoria. These cichlids still represent the fastest known rate of vertebrate speciation, whether they shared an ancestor 10,000 or 100,000 years ago. So, what can these phylogenetic studies tell us about the characteristics that predisposed the ancestor of the Lake Victoria flock to undergo such rapid radiation? Verheyen *et al.* suggest that

the immediate ancestors were already well adapted to the lake environment because *H. gracilior* shares the key ancestral haplotype with six other, ecologically diverse species from Lake Kivu. Likewise, Seehausen *et al.* point out that the genus *Thoracochromis* is already diverse in colour pattern and feeding habits, presaging the variation among species that occupy different ecological niches in Lake Victoria. But in fact the capacity for rapid and extensive radiation lay already in a *Haplochromis* species that lived at least a million years ago, because that species was the common ancestor of both the Lake Victoria flock and the spectacular parallel radiation of Lake Malawi cichlids in the southern Rift Valley⁸.

The challenge is to extend these phylogenies down to the level of incipient species. Only at this level can we study the selective forces acting on particular morphological and behavioural characteristics. Genomic techniques now enable us to identify the genes underlying key adaptive differences. It is the evolutionary history of these genomic regions in particular that should illuminate the selective mechanisms responsible for the fantastic diversity of East African cichlids. ■

Thomas D. Kocher is in the Department of Zoology, Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824, USA.
e-mail: Tom.Kocher@unh.edu

1. Johnson, T. C. *et al.* *Science* **273**, 1091–1093 (1996).
2. Verheyen, E., Salzburger, W., Snoeks, J. & Meyer, A. *Science* **300**, 325–329 (2003).
3. Nagl, S., Tichy, H., Mayer, W. E., Takahata, N. & Klein, J. *Proc. Natl Acad. Sci. USA* **95**, 14238–14243 (1998).
4. Nagl, S. *et al.* *Proc. R. Soc. Lond. B* **267**, 1049–1061 (2003).
5. Albertson, R. C., Markert, J. A., Danley, P. D. & Kocher, T. D. *Proc. Natl Acad. Sci. USA* **96**, 5107–5110 (1999).
6. Seehausen, O. *et al.* *Proc. R. Soc. Lond. B* **270**, 129–137 (2003).
7. Suzuki, Y., Glazko, G. V. & Nei, M. *Proc. Natl Acad. Sci. USA* **99**, 16138–16143 (2002).
8. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347**, 550–553 (1990).

Earth science

A slice of history

Paul D. Asimow

Investigations of an exposed slice of oceanic crust and mantle have provided a dramatic picture of temporal variation in the activity of the Mid-Atlantic Ridge — including its pulse rate of 3–4 million years.

In textbooks, mid-ocean ridges are often described as 'tape recorders': new crust is continuously being created by the partial melting of upwelling mantle and its eruption onto the ocean floor, where it solidifies and is 'rafted' away to either side of the ridge by seafloor spreading. The result — at least in textbooks — is an easily accessible map-view record of the formation of tectonic plates on the sea floor. The reality is much more messy. Not least, the tape is quickly obscured by sediments. In the central Atlantic, however,

at about 11° N, a quirk of movement has elevated a slice of sea floor and exposed a section through the crust and upper mantle at the Mid-Atlantic Ridge. This slice, known as the Vema transverse ridge, provides a view of 310 km, or 20 million years of Earth history. On page 499 of this issue¹, Bonatti *et al.* describe the insights into mid-ocean-ridge activity that may be gained from the Vema exposure.

To geophysicists using 'remote' techniques, the tape-recorder history remains



100 YEARS AGO

The extraction of the perfume from flowers such as jasmine, tuberose, violet and cassia has long been carried out by the process of enfleurage, the blossoms being left in contact with purified lard for a few days, and then replaced by fresh blossoms. The lard is either sold as such, or the essential oil may be extracted from it by melting it under strong alcohol. As the process of enfleurage is somewhat tedious, attempts have frequently been made to extract the oil directly from the flowers by means of light petroleum, but these processes have not as a rule proved successful, and it has recently been found that a very large proportion of the perfume is actually produced for the first time in the blossoms during the time occupied by the enfleurage. An interesting illustration of this is given by Dr. Albert Hesse in a recent number of the *Berichte*, in which he states that a ton (1000 kilos.) of tuberose blossoms only yielded 66 grams of oil when extracted with light petroleum, but during enfleurage yielded 801 grams of oil to the fat in which they were embedded, whilst a further 78 grams remained in the faded blossoms and could be separated by extraction or distillation.
From *Nature* 28 May 1903.

50 YEARS AGO

Although many papers have been written, there is still much that is not understood concerning the external and internal factors responsible for the cyclic nature of reproduction in reptiles. Investigations of the seasonal fluctuations in various parts of the reproductive system of several species of lizards, snakes and turtles have revealed interesting species differences. For a more complete understanding of reptilian reproduction it would appear that detailed descriptions of the seasonal changes of many species are desirable... Such a study was carried out by Wade Fox on the common garter snake (*Thamnophis*), several species and subspecies of which could easily be obtained throughout the year in the vicinity of San Francisco Bay. These investigations showed that the gonads of the male garter snake, *Thamnophis elegans terrestris*, exhibit marked seasonal variation. The testes begin to increase in size in the early spring, reach their maximum size and weight in late July, decline in the autumn, and are of minimum size during the winter months.
From *Nature* 30 May 1953.

readable for more than 100 million years in the magnetic, gravitational and topographical characteristics of ridges and their off-spring plates. But for those who depend on direct collection of rock samples for chemical analyses to determine the composition and origins of the rocks, the tape is ruined within tens of kilometres — equivalent to a few hundred thousand years — by sediment cover. A full picture of the processes occurring at mid-ocean ridges requires a '4-D' view of the architecture of the oceanic lithosphere, the cold and hence relatively rigid crust and upper mantle that constitute the tectonic plate. Yet from a chemical perspective, in most circumstances we can only work systematically in one dimension (along the ridge axis at zero age and at the seafloor depth at which the rocks are exposed).

The sequence of rock types to be found on the ocean floor is quite systematic, and each rock carries evidence of its origin in the crust-forming process. The upper mantle is dominated by rocks called peridotites. When plate spreading draws peridotite upwards beneath the ridge, the rock undergoes partial melting. The liquid fraction separates and rises to form the intrusive 'gabbroic' rocks of the lower oceanic crust and the erupted basalts that constitute the upper oceanic crust. The remaining solid material — the residual peridotite — is chemically modified by the loss of the more fusible elements (becoming richer, for example, in chromium relative to aluminium, in titanium relative to zirconium, and in magnesium relative to iron). Some of these residues of melting

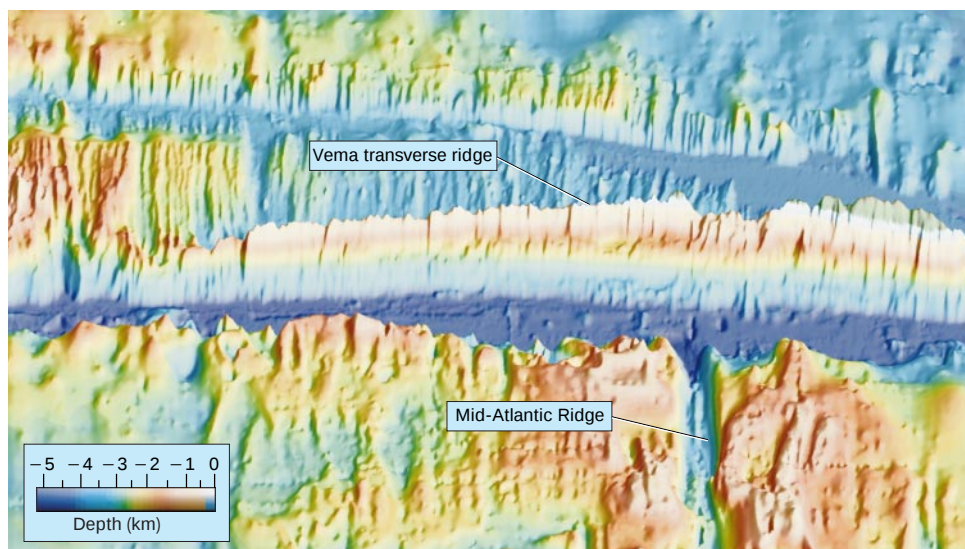


Figure 1 Ocean deep, mountain high. Under the waters of the central Atlantic, straddling the axis of the Mid-Atlantic Ridge, is a remarkable geological feature. The Vema transverse ridge is an exposed section of crust and mantle rocks, rising from a depth of more than 5,000 m at its valley floor to only 450 m below the sea surface at its highest point. Mapped by Bonatti *et al.*¹, and viewed here from the north, the ridge is a 20-million-year chronicle of tectonic activity.

eventually end up near the base of the crust and, as they cool, become incorporated into the lithosphere.

Rocks that have undergone melting, together with their residues, are evident elsewhere on Earth, of course — for instance in volcanoes on land. But attempts to use them to investigate Earth's interior are often confounded by the difficulty in distinguishing the source (the composition of the rock

that entered the melting region) from the process that modified it (the extent of melting, the temperatures and pressures involved, and the subsequent segregation, migration, differentiation, and so on). Yet at mid-ocean ridges, the entire crust is constructed in one episode of melting, and so the thickness of the crust provides a key overall constraint on the extent of melting. Given that there is a positive correlation

Astronomy

Sprinkler system

Henize 3-1475 is a planetary nebula — an expanding shell of gas surrounding a star that is entering the last stages of its life (and, in fact, nothing to do with planets at all). The gas shell of such an object may be spherical, dumb-bell-shaped or even completely irregular, but in the case of Henize 3-1475 it is concentrated into two fast-moving jets that emerge with velocities of up to 4 million kilometres per hour. As the system is also rotating, astronomers have nicknamed it the 'garden-sprinkler' nebula.

Angels Riera and colleagues have produced this composite image of Henize 3-1475 using different filters on a wide-field camera that is part of the Hubble Space Telescope. Coupled with observations from



ground-based telescopes, their data reveal that the fast outflow of gas from the central star is not regular. Instead, the gas is ejected in clumps, roughly once every hundred years. How the jets are created is still a mystery, but this clumping effect could be caused by interactions with a nearby companion star, or by magnetic processes in the central star (similar to the 22-year cycle of magnetic activity that creates sunspots on our own star).

The misnomer 'planetary nebula' was coined by William Herschel in 1784: cataloguing these objects, he was struck by how closely they resembled (viewed through a telescope considerably less powerful than Hubble) the planet he had recently discovered, Uranus.

Alison Wright

between crustal thickness and the mean extent of melting inferred from the chemistry of erupted basalts² and residual peridotites³, only the remaining, uncorrelated chemical variability can be attributed to the source.

This has led to the idea that variability along the mid-ocean-ridge system is mainly caused by temperature changes in the underlying mantle, which would affect the extent of melting of a reasonably homogeneous source. Such an idea is consistent with most (but not all) of the isotope data on the long-term history of the mantle sources of mid-ocean-ridge basalts. The zero-age view, however, gives little useful information about the geometry and dynamics of the flow of rock through the region of the mantle where melting occurs, or how this flow may vary with time.

This is where Bonatti and co-workers come in¹. The exposure of the Vema transverse ridge (Fig. 1) has allowed them to dredge up and analyse rock that would normally be deeply buried. They have therefore been able to measure an apparent time-delay between changes in crustal thickness, as estimated from gravity data, and in residual peridotite chemistry over the 20-million-year timescale. Their chain of logic derives from the following observations. The signals indicating crustal thickness and the extent of peridotite depletion show oscillations that have a wavelength of about 60 km (corresponding to a 3–4-million-year frequency), superposed on an overall 20-million-year trend of increasing peridotite depletion and crustal thickness. The variations seen in the two signals are correlated, but there is a phase lag of 22 km (or some 1.3 million years of spreading time) between the signals.

Bonatti *et al.* attribute the overall 20-million-year thickening trend to long-term warming of the mantle beneath the equatorial Atlantic, caused by southwards flow of mantle from 'hotspots' in the North Atlantic (although they acknowledge that it could also be due to along-axis growth of this particular ridge segment with time). The most exciting aspect of this study, though, is the interpretation of the 3–4-million-year signal. Experts might ask for a tighter correlation between the time series of crustal thickness and extent of melting derived from peridotites, but this correlation seems to dispense with the need to invoke explanations for the observed variation in crustal thickness that depend entirely on melt migration or variation in the composition of the source rock.

That leaves variations in mantle temperature and flow geometry. There is no obvious reason why mantle temperature should oscillate at this 3–4-million-year frequency, but dynamical calculations, making certain assumptions about viscosity changes and the internal buoyancy of basalts and residual

peridotites, predict that there would be bursts of increased flow through the melting region⁴. These episodes of rapid flow can explain times of high crustal thickness. But the mechanism by which they generate more depleted residual peridotites is unclear, as is the basis of the sawtoothed form of the crustal-thickness signal (as strikingly depicted in Fig. 3d on page 501). Bonatti *et al.* propose that the flow variation is restricted to the deeper, water-bearing, and hence low-viscosity, part of the melting regime⁵ and that the intervals of active flow somehow deliver hotter material to the shallower, high-viscosity part of the melting regime. An alternative explanation is that the episodes of active flow change the balance between advection and conduction in the shallowest part of the melting region, as hinted at by equilibration temperatures that record different cooling rates in the same peridotites.

What of the 22-km offset between the signals denoting residual peridotite and crustal thickness? Melt migration is much more rapid than solid flow through the melting region, so there would be a time delay between melt extraction from a given parcel of peridotite to create basalts, and the eventual emplacement of that same peridotite as a residue at the base of the lithosphere. In other words, the oscillation signals carried by the melt volume and by peridotite chemistry propagate upwards at different speeds and are recorded at different times, being spatially offset by the plate spreading rate. Bonatti *et al.* use the phase lag between their extent-of-melting signals from peridotite and from crustal thickness (assuming that

the bulk of the melt separates at a depth of 35 km and that melt velocity can be taken as infinite) to calculate the solid upwelling rate, and they arrive at an estimate of 25 mm per year. This is faster than the rate at which the plate is moving away from the ridge (14–17 mm per year) and hence is consistent with a component of buoyant flow.

Mid-ocean ridges have different characteristics according to their speed of spreading. At the Vema anomaly, the Mid-Atlantic Ridge is slow spreading, and Bonatti and colleagues' study adds to our understanding of how the oceanic crust grows in these circumstances. Similar understanding of fast-spreading sections, which do not tend to create such spectacular tectonic exposures, seems further away, but there is plenty more to learn from the Vema feature. For instance, it will be interesting to see whether informative chemical variations have been preserved in the basaltic rocks at the top of the section, and whether they correlate with the picture that Bonatti *et al.* have compiled from the data on crustal thickness and exposed mantle rocks. ■

Paul D. Asimow is in the Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Boulevard, Pasadena, California 91125, USA.
e-mail: asimow@gps.caltech.edu

1. Bonatti, E. *et al.* *Nature* **423**, 499–505 (2003).
2. Klein, E. M. & Langmuir, C. H. *J. Geophys. Res.* **92**, 8089–8115 (1987).
3. Dick, H. J. B., Fisher, R. L. & Bryan, W. B. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
4. Scott, D. R. & Stevenson, D. J. *J. Geophys. Res.* **94**, 2973–2988 (1989).
5. Braun, M. G., Hirth, G. & Parmentier, E. M. *Earth Planet. Sci. Lett.* **176**, 339–356 (2000).

Cell division

Genome maintenance

Jordan W. Raff

Early fruitfly embryos have an unusual means of halting the division of any nuclei containing damaged DNA. A key component of this mechanism has now been identified, and might have implications for cancer.

Maintaining the integrity of the genome is a crucial task for any cell. Two proteins, called checkpoint kinases 1 (Chk1) and 2 (Chk2), help to achieve this in many species, and mutations in the genes encoding these proteins have been linked to the generation of cancer in humans. The proteins are activated by DNA damage, and help to initiate cellular defence responses that include the stimulation of DNA-repair pathways and the slowing down of the cell-division cycle to allow time for repair^{1,2}. In multicellular organisms, if the DNA is not successfully mended, the damaged cells usually kill themselves — thereby eliminating the defective genome.

As they describe in *Cell*, Theurkauf and colleagues³ have discovered that Chk2 is also involved in a rather different defence mechanism that is triggered by DNA damage in early fruitfly embryos.

This particular defence response is especially well suited to the early fruitfly (*Drosophila*) embryo, in which the cell nuclei undergo a series of 13 rapid divisions within a common cytoplasm⁴. These swift nuclear divisions occur synchronously, and consist entirely of alternating phases of DNA synthesis (S phase) and DNA segregation (mitosis or M phase), with none of the intervening gap phases that separate S and M in more typical cell cycles. Because there is no gap

phase, if one of the embryonic nuclei is damaged during S phase, it does not have the usual option of stopping the division cycle before mitosis. Instead, it will be driven into mitosis, in synchrony with the surrounding undamaged nuclei in the common cytoplasm.

During nuclear-division cycles 9 and 10, most nuclei migrate to the outer rim of the embryo — the cortex (Fig. 1a). A few remain in the interior, but these will not contribute to the adult fly. It has been known for some time that, during cycles 10 to 13, any cortical nucleus that suffers DNA damage eventually drops into the interior of the embryo and is thereby effectively eliminated from the organism⁵. It has been a mystery, however, how such damaged nuclei are recognized and how they are then discharged into the interior.

Three years ago, Theurkauf and colleagues⁶ described the phenomenon of centrosome inactivation in *Drosophila* embryos. Centrosomes are structures that are needed to efficiently segregate DNA during mitosis. They contain so-called γ -tubulin ring complexes (γ -TURCs), which organize the long filaments, or microtubules, that make up the mitotic spindle — a bipolar apparatus on which chromosomes are segregated. Theurkauf and co-workers⁶ noticed that nuclei that failed to complete DNA synthesis or suffered DNA damage during S phase formed abnormal spindles when they entered mitosis. This was apparently because the DNA damage triggered the displacement of the γ -TURCs from the centrosomes during mitosis. Intriguingly, the γ -TURCs reappeared at centrosomes after mitosis was complete, and several other centrosomal proteins remained concentrated at centrosomes during mitosis, hinting that a core centrosome structure remained intact throughout the division cycle. The abnormal nuclei that reformed after the aberrant mitosis then rapidly dropped into the interior of the embryo, and so were effectively eliminated (Fig. 1b–d).

Theurkauf and colleagues³ have now shown that double-stranded DNA breaks are responsible for triggering this centrosome inactivation, and that Chk2 is essential for the process. In vertebrate cells, many Chk2-dependent responses to DNA damage are induced via the activation of the p53 protein, but the authors found that this is not the case for centrosome inactivation. Moreover, they discovered that Chk2 itself becomes concentrated at centrosomes, and that DNA damage seems to enhance its accumulation there. So, given that Chk2 is a kinase — it modifies proteins by phosphorylating them — perhaps it inactivates centrosomes by directly phosphorylating one or more of their protein components.

Might Chk2 also induce centrosome inactivation in other cell types? As mentioned

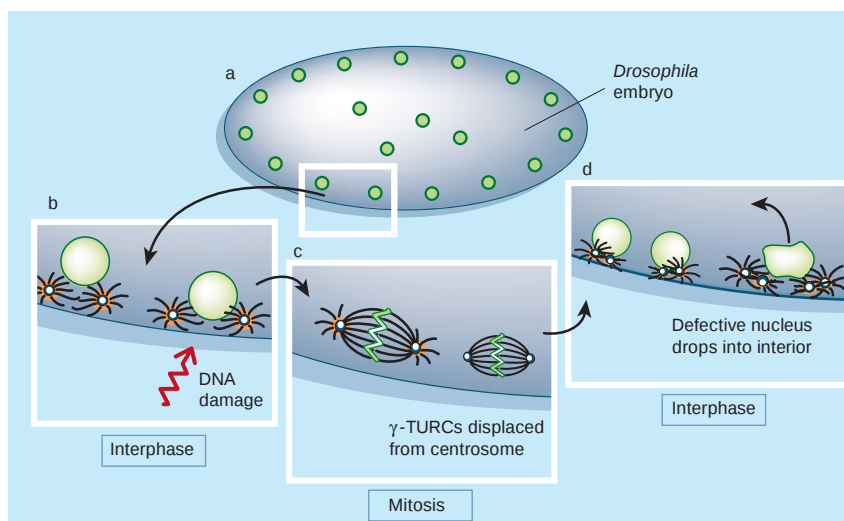


Figure 1 Dealing with DNA damage in *Drosophila*. a, An early *Drosophila* embryo that has undergone 10–13 nuclear divisions. Most nuclei (green) are aligned around the outside, with only a few in the interior. These internal nuclei — ‘yolk nuclei’ — will not contribute to adult tissues. b–d, The fate of a nucleus whose DNA has been damaged in interphase (the period between one nuclear division and the next). b, At the time of DNA damage the centrosome is unaffected, and microtubules (black) are still organized by γ -tubulin ring complexes (γ -TURCs; orange) that are concentrated around a core centrosomal structure. c, As all the undamaged nuclei enter mitosis, so too does the damaged nucleus, but the γ -TURCs appear to be released from the centrosomes and the mitotic spindle (black) does not form properly. (Note that, at this point, the nuclear membrane has disintegrated, allowing the spindle to attach to chromosomes.) Chromosome segregation fails and, in the following interphase (d), the defective nucleus falls into the interior of the embryo, leaving the centrosomes behind. Theurkauf and colleagues³ have found that the Chk2 protein is crucial for this nuclear defence mechanism.

above, in most cells, DNA damage that occurs in the phases between mitoses (these phases are collectively known as interphase) causes a cell-cycle arrest prior to mitosis^{1,2}. But this cannot occur in early *Drosophila* embryos, so a unique Chk2-dependent mechanism may have evolved to eliminate defective nuclei during mitosis instead. In support of this possibility, DNA damage in older, cellularized fruitfly embryos (when the nuclei no longer share a common cytoplasm, and there is a gap phase between S and M phases) does not appear to lead to centrosome inactivation, but instead causes a delay in both the entry into and exit from mitosis^{7,8}. And in vertebrate cells in culture, DNA damage during mitosis does not induce centrosome inactivation⁹ — although, even in early *Drosophila* embryos, DNA damage that occurs during mitosis does not appear to trigger centrosome inactivation, implying that Chk2 may need to be activated in interphase to cause centrosome inactivation in mitosis^{3,6}.

It seems unlikely, then, that centrosome inactivation is a major response to DNA damage in most normal cells, as DNA damage during interphase will usually lead to cell-cycle arrest before the cell enters mitosis. If, however, a cell manages to enter mitosis with unrepaired DNA damage, it might then become important to trigger centrosome inactivation in order to eliminate

the defective DNA. In fact, there is some evidence that vertebrate cells that enter mitosis carrying damaged DNA die by a ‘mitotic catastrophe’¹⁰, although it is not clear whether this mechanism requires Chk2 or is indeed caused by centrosome inactivation. If such a process does exist, however, it might help to protect against cancer. The mechanisms that monitor DNA damage are often impaired in pre-cancerous cells, and so it may be relatively common for such cells to enter mitosis carrying defective DNA. If centrosome inactivation proves to be more than just a specialization of flies, the race will be on to understand how Chk2 brings it about, and to test whether it is involved in preventing human cancer.

Jordan W. Raff is at The Wellcome Trust/Cancer Research UK Institute, Tennis Court Road, Cambridge CB2 1QR, UK.
e-mail: j.raff@welc.cam.ac.uk

- Dasika, G. K. et al. *Oncogene* **18**, 7883–7899 (1999).
- Taylor, W. R. & Stark, G. R. *Oncogene* **20**, 1803–1815 (2001).
- Takada, S., Kelkar, A. & Theurkauf, W. E. *Cell* **113**, 87–99 (2003).
- Foe, V. E. & Alberts, B. M. *J. Cell Sci.* **61**, 31–70 (1983).
- Sullivan, W., Daily, D. R., Fogarty, P., Yook, K. J. & Pimpinelli, S. *Mol. Biol. Cell* **4**, 885–896 (1993).
- Sibon, O. C., Kelkar, A., Lemstra, W. & Theurkauf, W. E. *Nature Cell Biol.* **2**, 90–95 (2000).
- Su, T. T., Walker, J. & Stumpff, J. *Curr. Biol.* **10**, 119–126 (2000).
- Su, T. T. & Jaklevic, B. *Curr. Biol.* **11**, 8–17 (2001).
- Mikhailov, A., Cole, R. W. & Rieder, C. L. *Curr. Biol.* **12**, 1797–1806 (2002).
- Roninson, I. B., Broude, E. V. & Chang, B. D. *Drug Resist. Updates* **4**, 303–313 (2001).

Microfluidics

DNA microarrays
get the shakes*Anal. Chem.* **75**, 1911–1917 (2003)

At the microscopic scale, water acts like a highly viscous liquid and mixing is slow. This is a problem for microfluidic chemistry, as it slows down the rate of encounters between reagents. Genetic analysis of DNA on microchip arrays, which relies on target and probe strands of DNA getting together, can take hours because this hybridization process is limited by the rate of strand diffusion.

Robin Hui Liu and colleagues have shown that hybridization can be speeded up by using the acoustic vibration of tiny air bubbles to stir the liquid. They position a plastic film just above a glass microarray chip, perforated with tiny holes that each contains an air bubble. A piezoelectric oscillator makes the bubbles vibrate, creating a 'microstreaming' circulation of the liquid between the plastic film and the glass slide. This speeds up hybridization roughly fivefold, so that the microarray read-out signal is correspondingly stronger after two hours. And the resulting signal is much more uniform over the chip. This simple, cheap technique could be a boon to high-throughput microarray-based genetic screening.

Philip Ball

Sensory transduction

Feel the heat

Science **300**, 1284–1288 (2003)

Bite into a chilli and you'll soon know about it. The burning sensation that follows is caused by capsaicin, a compound that stimulates receptors found on the surface of sensory nerve cells in the tongue and elsewhere. These so-called TRPV1 receptors form part of a warning system, signalling the presence of various potentially harmful stimuli, including low pH and noxious heat. Elizabeth D. Prescott and David Julius now show how the sensitivity of TRPV1 receptors is regulated to ensure that we detect and respond to such threats appropriately.

The researchers looked at the interaction of the TRPV1 protein with the lipid molecule PIP₂, which can hold the receptor in an inhibited state. By modifying the structure of TRPV1 in various ways, the authors identified the site at which PIP₂ binds. Structural changes that weakened the interaction of TRPV1 with PIP₂ increased the response of the receptor to chemical and thermal triggers, reducing the threshold for detecting these stimuli.

Prescott and Julius suggest that the strength of the TRPV1–PIP₂ interaction sets the threshold for detecting the onset of



tissue damage. They also propose that the PIP₂-binding site represents a target for treatments aimed at altering the sensitivity of nerve cells to damaging stimuli in normal and pathological conditions.

Rebecca Craven

Earth science

Core effect

Geology **31**, 415–418 (2003)

Life on Earth 2.45 billion years ago was harsh, but it was safely swaddled in the same protective magnetic field that we enjoy today, according to A. V. Smirnov and colleagues.

The magnetic field is generated by the dynamo effect of molten iron sloshing around in the Earth's core; by repelling charged particles, the magnetic field shields the planet from harsh radiation from space. But it's difficult to know when the Earth's dynamo turned on. The best method for finding out is to measure the magnetic field remaining in rocks to infer the strength of the magnetic field that ran through them in days past. But this picture is often clouded by the weathering and mixing of rocks over time.

Smirnov *et al.* have measured the magnetic field in 2.45-billion-year-old rocks from Karelia in Russia. To overcome the problem of weathering, they looked at individual crystals of rock — the first time this has been done with rocks of this vintage. They find that the magnetic field present 2.45 billion years ago was similar to that of today. This implies that the Earth's solid iron inner core — believed to stabilize the dynamo — had started to form before that time. If correct, this is an important yardstick for Earth researchers.

Tom Clarke

Cell biology

Rafting for dear life

Immunity **18**, 655–664 (2003)

Cell membranes are made up of a sea of lipid molecules, in which so-called rafts are afloat — these are membrane patches that are high in sphingolipids and cholesterol. Rafts can serve as launching pads for transmitting signals into the cell. Daniel F. Legler *et al.* now report that, when activated, a receptor for the tumour-necrosis factor- α (TNF- α)

protein is recruited to rafts, where it launches a molecular signalling cascade necessary for the cell to survive.

Usually, the activation of this receptor (TNF receptor 1) by TNF- α leads to the production of anti-inflammatory and survival factors — but occasionally it results in programmed cell death. Legler and colleagues found that, following activation and adrift on its raft, the receptor attracted various signalling proteins, which formed a complex that could transmit the signal through the cell. The receptor and one of the proteins in the complex also rapidly became modified — most probably with ubiquitin molecules. These molecules label proteins for rapid destruction and may be necessary to quench the signal at some point.

Most importantly, disrupting the organization of the rafts switched TNF- α -induced signalling from cell survival to cell death. The role of rafts in TNF- α signalling is not always clear-cut, but here it looks as if the cells have to raft for dear life to survive.

Marie-Thérèse Heemels

Cancer

Genetic prescription

Clin. Cancer Res. **9**, 1611–1615 (2003)

A patient's genetic make-up may affect how well they respond to anti-cancer treatments. According to Victor Cohen *et al.*, people with colorectal cancer respond better to a common type of chemotherapy if they have a particular amino-acid sequence variation in the enzyme methylenetetrahydrofolate reductase (MTHFR), which is involved in folic acid metabolism.

Colorectal cancer is commonly treated with fluoropyrimidine compounds, which affect the composition of the pool of folic acid derivatives inside cells. This pool is also affected by a sequence variation in MTHFR (where an alanine amino acid is changed to a valine), which reduces the activity of the enzyme. It has been suggested that this variation is associated with a decreased risk of colorectal cancer.

Cohen and colleagues wondered whether the variation also correlates with a better response to fluoropyrimidines. They carried out an initial study of 43 patients with colorectal cancer, 26 of whom carried the valine variant of MTHFR. Tumours were less likely to develop further in patients with this variant than in patients with the alanine variant.

The findings point to 'personalized' treatments, and to a possible new drug target. But first a larger-scale, randomized study is needed to find out how accurately the sequence variation predicts the response to treatment and whether it has other long-term consequences.

Camilla Carlsson

Violence viewed by psychopathic murderers

Adapting a revealing test may expose those psychopaths who are most likely to kill.

Psychopathic murderers are often portrayed as cold-blooded, emotionless and lacking in remorse¹, but they are also adept at lying and at feigning the emotions in which they are deficient. Here we adapt a test known as the Implicit Association Test (IAT)², which was previously used to assess concealed prejudices, to show that psychopathic murderers have abnormal cognitive associations regarding violence, which may underpin their actions. Such implicit measures may provide us with an important insight into the criminal mind.

The ability of psychopaths to con, lie and manipulate means that their potential danger to others may not be appreciated by criminal-justice agencies, that their response to clinical assessment is often rigged³, and that researchers are blinded to underlying motives (Fig. 1). Faced with the problem of detecting socially stigmatic beliefs, psychologists have developed tasks that measure these attitudes implicitly. The IAT has successfully quantified beliefs that people may wish to disguise, such as negative views about racial groups, homosexuality or obesity. We have adapted the IAT to reveal implicit beliefs about violence in psychopathic murderers.

We followed the general methodology of the IAT². Briefly, uppercase words (for example, 'UGLY') are classified as being 'pleasant' or 'unpleasant', and lowercase words (for example, 'kill') are classified as 'violent' or 'peaceful', by pressing corresponding buttons. When the same response key is assigned for both the unpleasant and

violent words (this is termed the congruent condition), most people find the task easy. But when pleasant and violent words share the same response key (the incongruent condition), most people find this confusing. The association between 'pleasant-unpleasant' and 'violent-peaceful' is indexed by means of the IAT effect (reaction time for the incongruent condition minus reaction time for the congruent condition).

Our sample ($n=121$) was recruited from consecutive admissions to a secure therapeutic community for male offenders with personality disorders. None of the participants had a co-morbid mental illness and none was taking any illicit drug or medication. Psychopathy was measured using the Psychopathy Checklist-Revised test⁴; those who scored over 25 were labelled as psychopaths⁵. We selected the murderers in order to target the most seriously violent group. Our final groups consisted of 13 psychopathic murderers, 17 non-psychopathic murderers, 39 psychopathic other offenders and 52 non-psychopathic other offenders. There was no significant difference in intelligence quotient (IQ) among the groups.

The results from the 'violent' IAT test are shown in Fig. 2. Statistical analysis (analysis of variance) indicates that there was no significant effect due to grouping (murderer versus other offender) or psychopathy score. Crucially, the interaction between these variables was significant ($P<0.05$) and remained so even when age and IQ (as assessed by the revised National Adult Reading Test⁶) were added as covariates. As expected, the psychopathic murderers showed a much lower IAT effect than the non-psychopathic murderers ($P<0.05$). There was no significant difference in error rate related to offence type, psychopathy or interaction between these variables (Fig. 2, inset). A control IAT (flower-insect; pleasant-unpleasant) revealed no effect for psychopathy or offence type.

Our results indicate that the reduced violent-IAT effect seen in psychopathic murderers is likely to be due to their abnormal beliefs about violence, rather than to some other nonspecific effect such as poor impulse control and/or deficits in decision-making. Psychopathic murderers have diminished negative reactions to violence compared with non-psychopathic murderers and other offenders.

To our knowledge, this is the first cognitive demonstration of abnormal social beliefs in murderous psychopaths. The interaction between psychopathy and murder suggests that it is the combination of these two factors



Figure 1 Portrait of a serial killer: psychopathic murderers such as Ted Bundy can often project a deceptively charming persona.

AP

that is crucial in identifying deficient implicit beliefs about violence, resulting in a relatively rare subpopulation of very violent psychopaths. Obviously, not all murderers have similar motivations or beliefs about violence, and not all psychopaths commit murder (as illustrated by the phenomenon of the 'white-collar' psychopath⁷).

Our results have some testable implications. For example, non-psychopaths could become sensitized to violence after committing murder⁸ (hence their greater scores in the IAT). The difference between the psychopathic groups may reflect two separate, stable populations of psychopathic offenders: one with deficient social beliefs (and therefore an increased disposition towards extreme violence) and the other in which such negative beliefs are absent. If this difference can be picked up by the 'violent' IAT before an offence is committed, this test may become an important tool for distinguishing psychopaths who are likely to commit extremely violent offences from those who are not.

Nicola S. Gray*†, Malcolm J. MacCulloch*†‡, Jennifer Smith*, Mark Morris§, Robert J. Snowden*

*School of Psychology, Cardiff University, Cardiff CF10 3YG, UK

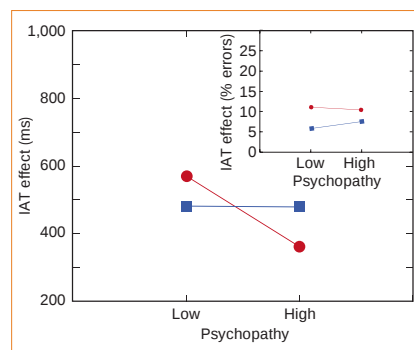


Figure 2 Results of the Implicit Association Test (IAT) effect (difference in reaction time between incongruent and congruent conditions; see text), plotted as a function of psychopathy score (PCL-R⁴) for murderers (red) and non-murderers (blue). Inset, IAT effect for error rates. All participants gave written informed consent. We used eight 'pleasant' and eight 'unpleasant' words⁹, and eight 'violent' and eight 'peaceful' words. Both the congruent and incongruent blocks consisted of 64 trials, with words presented in a pseudo-random order. Further details are available from the authors.

e-mail: grayns@cardiff.ac.uk

†South Wales Forensic Psychiatric Service,
Caswell Clinic, Glanrhyd Hospital,
Mid Glamorgan CF36 4LN, UK

‡Department of Psychological Medicine,
University of Wales College of Medicine, Heath
Park, Cardiff CF14 4XN, UK

§HM Prisons, Grendon and Springhill,
Grendon Underwood, Aylesbury,
Buckinghamshire HP18 0TL, UK

1. Cleckley, H. *The Mask of Sanity* (Mosby, St Louis, 1941).

Prion diseases

BSE in sheep bred for resistance to infection

Selective breeding for disease-resistant genotypes is being pursued as a means of eradicating scrapie (and bovine spongiform encephalopathy (BSE), if it is present) from sheep flocks. Here we show that the genotype associated with the highest resistance can still be infected with BSE by intracerebral inoculation. Although the relevance of this finding to sheep exposed to natural infection remains to be determined, it may have important implications for disease-eradication strategies.

Susceptibility and resistance to transmissible spongiform encephalopathies (TSEs) in sheep is largely controlled by polymorphisms at codons 136, 154 and 171 of the gene that encodes prion protein (PrP). Representing the resulting alleles as the amino acids at these positions¹, susceptibility to scrapie is strongly associated with valine (V) at position 136 and/or glutamine (Q) at position 171, and susceptibility to experimental BSE is associated predominantly with the alanine-arginine-glutamine (ARQ) allele².

The ARR allele is associated with resistance to natural scrapie — the disease is rare in ARR heterozygotes³ and only one unconfirmed case has ever been reported in a homozygote⁴. The limited data available suggest that ARR/ARR sheep are also resistant to experimental oral challenge with scrapie and BSE^{5,6}. Britain has therefore initiated the National Scrapie Plan, which aims eventually to eradicate scrapie (as well as any BSE concealed in sheep)⁷ by means of increasing the frequency of the ARR allele and decreasing that of the VRQ allele.

We investigated whether ARR/ARR sheep are still resistant to transmission of BSE when inoculated intracerebrally with infected cattle-brain homogenate (0.05 g). Groups of New Zealand (scrapie-free) sheep of three different breeds (Suffolk, Cheviot and Poll Dorset) and six different PrP genotypes (VRQ/VRQ, VRQ/ARQ, VRQ/ARR, ARQ/ARQ, ARQ/ARR and ARR/ARR) were tested.

- Greenwald, A. G., McGhee, J. L. & Schwartz, J. L. *J. Person. Soc. Psychol.* **74**, 1464–1480 (1998).
- Rice, M. E., Harris, G. T. & Cormier, C. A. *Law Hum. Behav.* **16**, 399–412 (1992).
- Hare, R. D. *The Hare Psychopathy Checklist Revised* (Multi-Health Systems, Toronto, 1991).
- Cooke, D. L. & Michie, C. J. *Abnorm. Psychol.* **108**, 58–68 (1999).
- Nelson, H. *National Adult Reading Test (NART) Manual* (NFER-Nelson, Windsor, 1982).
- Hare, R. D. *Without Conscience* (Guilford, New York, 1999).
- Gray, N. S. et al. *J. Forens. Psychiat. Psychol.* **14**, 27–43 (2003).
- Bellezza, F. S., Greenwald, A. G. & Banaji, M. R. *Behav. Res. Methods Instr. Comp.* **18**, 299–303 (1986).

Competing financial interests: declared none.

As expected, the incubation period (495–671 days) was shortest in ARQ/ARQ sheep, with 17 out of 19 being affected. So far, four out of ten VRQ/VRQ and two out of ten VRQ/ARQ Cheviot sheep have also succumbed to disease after incubation periods of 881–1,092 days. Unexpectedly, 3 out of 19 BSE-challenged ARR/ARR sheep also showed clinical symptoms after incubation periods of 1,008, 1,124 and 1,127 days, respectively, but none of the ARR heterozygotes has been clinically affected to date.

Western-blot analysis of the abnormal protease-resistant fraction of PrP (PrP^{Sc}) (results not shown) and immunohistochemical labelling of disease-associated PrP in brain sections from the ARR/ARR sheep (Fig. 1) were consistent with experimental BSE in ARQ/ARQ sheep^{8,9}. This contrasts with the results of *in vitro* cell-free conversion experiments, in which conversion of the sheep ARR PrP protein to PrP^{Sc} was extremely inefficient¹⁰.

The susceptibility of ARR/ARR sheep to intracerebral injection with BSE indicates that these animals cannot be regarded as having absolute genetic resistance to TSE infection. The significance of this result for sheep that are exposed to natural TSE infection remains to be determined. In ongoing experiments, ARR/ARR sheep orally dosed with BSE have remained free from clinical disease or evidence of preclinical infection for up to 2,100 days after BSE challenge (ref. 6

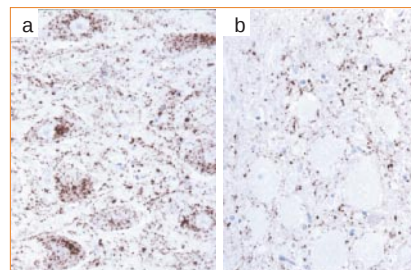


Figure 1 Immunohistochemical labelling of abnormal PrP protein in the dorsal motor nucleus of the vagus of a sheep that is homozygous for the ARR allele (see text) and has been infected with bovine spongiform encephalopathy (BSE). **a**, Intraneuronal and neuropil-associated accumulations are revealed by the R145 antibody; **b**, the antibody P4 reveals neuropil-associated (but not intraneuronal) extracellular accumulation. This pattern of labelling can distinguish BSE from scrapie in sheep.

and J.F., unpublished results). However, the number of animals in these studies is probably too small to rule out susceptibility even if the surviving animals remain free of infection.

Our results may reflect those found for pigs, in which no natural case of BSE has been reported. Pigs inoculated with BSE simultaneously by intracerebral and peripheral routes developed TSE within 520–1,130 days, whereas those exposed to repeated oral challenge remained healthy for up to 2,550 days after infection¹¹.

Even if ARR/ARR sheep exposed to natural TSE infection do not succumb to disease, they might be able to act as sub-clinical carriers of infection. However, no abnormal PrP (and, by implication, infectivity) was detected in the lymphoid tissues of any of the ARR/ARR sheep cases of BSE reported here. Likewise, neither clinically normal ARR/ARR sheep in scrapie-endemic flocks, nor those challenged orally with TSEs, show evidence of abnormal PrP deposition in peripheral tissues — although here again the number of animals examined is small^{6,12}. These results indicate that ARR/ARR sheep are unlikely to excrete the infectious agent, although transmission of infection to in-contact animals in the absence of detectable abnormal PrP cannot be ruled out.

We have shown that ARR/ARR sheep can become infected when inoculated intracerebrally with a high dose of BSE, but there is, as yet, no evidence to suggest that they are susceptible when exposed by natural routes of infection. But the fact that the ARR PrP protein can undergo conversion to a pathological isoform *in vivo* raises questions about the basic mechanisms that underpin genetic resistance to TSEs.

Fiona Houston*, **Wilfred Goldmann†**,
Angela Chong†, **Martin Jeffrey‡**,
Lorenzo González‡, **James Foster†**,
David Parnham†, **Nora Hunter†**

*Institute for Animal Health, Compton, Newbury,
Berkshire RG20 7NN, UK

e-mail: fiona.houston@bbsrc.sc.uk

†IAH Neuropathogenesis Unit, Ogston Building,
Edinburgh EH9 3JF, UK

‡Veterinary Laboratories Agency Lasswade,
Pentlands Science Park, Penicuik,
Midlothian EH26 0PZ, UK

- Hunter, N. *Trends Microbiol.* **5**, 331–334 (1997).
- Goldmann, W., Hunter, N., Smith, G., Foster, J. & Hope, I. *J. Gen. Virol.* **75**, 989–995 (1994).
- van Keulen, L. J. et al. *J. Clin. Microbiol.* **34**, 1228–1231 (1996).
- Ikedu, T. et al. *J. Gen. Virol.* **76**, 2577–2581 (1995).
- O'Rourke, K. I. et al. *J. Gen. Virol.* **78**, 975–978 (1997).
- Jeffrey, M. et al. *J. Comp. Pathol.* **124**, 280–289 (2001).
- Kao, R. R. et al. *Science* **295**, 332–335 (2002).
- Hope, J. et al. *J. Gen. Virol.* **81**, 851 (2000).
- Jeffrey, M. et al. *J. Comp. Pathol.* **125**, 271–284 (2001).
- Bossers, A., de Vries, R. & Smits, M. A. *J. Virol.* **74**, 1407–1414 (2000).
- Ryder, S. J., Hawkins, S. A., Dawson, M. & Wells, G. A. *J. Comp. Pathol.* **122**, 131–143 (2000).
- Jeffrey, M. et al. *J. Comp. Pathol.* **127**, 264–273 (2002).

Competing financial interests: declared none.

Mantle thermal pulses below the Mid-Atlantic Ridge and temporal variations in the formation of oceanic lithosphere

Enrico Bonatti^{*†‡}, Marco Ligi^{*}, Daniele Brunelli^{*†}, Anna Cipriani[‡], Paola Fabretti^{*}, Valentina Ferrante^{*†}, Luca Gasperini^{*} & Luisa Ottolini[§]

^{*} Istituto di Scienze Marine, Geologia Marina, CNR, Via Gobetti 101, 40129, Bologna, Italy

[†] Dipartimento di Scienze della Terra, Università "La Sapienza", Piazzale Aldo Moro 5, 00187, Rome, Italy

[‡] Department of Earth and Environmental Sciences, Lamont Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

[§] Istituto di Geoscienze e Georisorse, Sezione di Pavia, CNR, Via Ferrata 1, 27100, Pavia, Italy

A 20-Myr record of creation of oceanic lithosphere at a segment of the central Mid-Atlantic-Ridge is exposed along an uplifted sliver of lithosphere. The degree of melting of the mantle that is upwelling below the ridge, estimated from the chemistry of the exposed mantle rocks, as well as crustal thickness inferred from gravity measurements, show oscillations of ~3–4 Myr superimposed on a longer-term steady increase with time. The time lag between oscillations of mantle melting and crustal thickness indicates that the solid mantle is upwelling at an average rate of ~25 mm yr⁻¹, but this appears to vary through time. Slow-spreading lithosphere seems to form through dynamic pulses of mantle upwelling and melting, leading not only to along-axis segmentation but also to across-axis structural variability. Also, the central Mid-Atlantic Ridge appears to have become steadily hotter over the past 20 Myr, possibly owing to north–south mantle flow.

The oceanic lithosphere covers two-thirds of our planet: understanding how it forms and evolves is a major challenge in the Earth sciences. It is generally agreed that the oceanic lithosphere forms along mid-ocean ridges, where mantle material upwells and undergoes decompression and partial melting. The melt rises rapidly and freezes, producing the crust, while the peridotitic residue forms the lithospheric mantle. Mid-ocean-ridge topography, structure and composition indicate that near zero age processes of lithosphere formation vary along ridge axis^{1–3}. Less is known, however, of how these processes vary through time, a question important for our understanding of how ocean basins evolve. Variations through time of the thermal regime and/or composition of a mid-ocean ridge should be recorded in lithosphere lying at increasing distances from the ridge axis along sea-floor spreading flow lines. However, older lithosphere is normally covered by sediment and not easily accessible to high-resolution observation and sampling.

An uplifted sliver of oceanic lithosphere (Fig. 1), exposing in the central Atlantic an ~20-Myr-long record of creation of lithosphere at a segment of the Mid-Atlantic Ridge (MAR), gave us the opportunity to investigate the temporal variability in the formation of lithosphere, to estimate the upwelling velocity of the mantle below the ridge, and to determine whether passive or dynamic models of creation of oceanic lithosphere prevail at slow-spreading ridges.

A sliver of exposed oceanic lithosphere

A major topographic anomaly, the Vema transverse ridge, runs on the southern side of the Vema transform, which offsets the MAR by 310 km at ~11° N (Fig. 1). Its exposed northern scarp (that is, the southern wall of the transform valley) reaches in height up to 4 km above the valley floor (Fig. 2). Submersible⁴, multibeam⁵ and seismic reflection data, and bottom samples (Fig. 1) revealed that the northern side of the Vema transverse ridge exposes a ~3–4-km-thick, relatively complete and tectonically undeformed upper

lithospheric section (Vema lithospheric section, VLS). The VLS consists of a ~1 km mantle peridotite basal unit, overlain by a gabbroic unit, then by a dyke complex and topped by pillow basalt. In contrast, the southern slope of the transverse ridge exposes only upper crustal basalt (Fig. 2). The VSL is the edge of a slab of oceanic lithosphere that was flexured and uplifted between 10 and 12 Myr ago⁶, and is now exposed continuously for at least 300 km along a sea-floor-spreading flow-line. This corresponds to a time interval of roughly 20 Myr, if we assume that this lithosphere formed at the eastern MAR segment (EMARS) that meets the Vema fracture zone from the south (Fig. 1), with no ridge jumps during this interval. Spreading half rates, estimated from plate motion reconstructions⁷, are 13.6 mm yr⁻¹ between 0 and 10 Myr ago, 16.9 mm yr⁻¹ between 10 and 19 Myr ago, and 17.2 mm yr⁻¹ between 19 and 26 Myr ago. Crustal ages cited here are estimated assuming these rates.

We investigated lateral (that is, temporal) variations of lithospheric thermal structure and composition along the VLS through two independent approaches: (1) geochemistry of rocks obtained by systematic sampling; and (2) satellite and shipboard gravimetry.

Temporal variations of degree of melting

We sampled systematically the lower unit of the VLS at a maximum horizontal spacing of ~10 km, along a stretch of 250 km, that is, ~15 Myr (Figs 1 and 3). We obtained ultramafic rocks at all sites below 4 km depth, that is, within the lower 1 km portion of the section. Serpentinized protogranular/porphyroclastic peridotites prevail; they contain olivine, orthopyroxene (opx), clinopyroxene (cpx) and spinel (sp) as relict mantle-equilibrated minerals. They are generally similar to ultramafics sampled elsewhere along mid-ocean ridges, considered to be the residue from various extents of partial melting that took place below the ridge axis^{8–10}.

We determined by electron probe the major-element composition of mantle-equilibrated phases opx, cpx and spinel from more than 50 ultramafic rocks. Opx and cpx were also analysed for trace

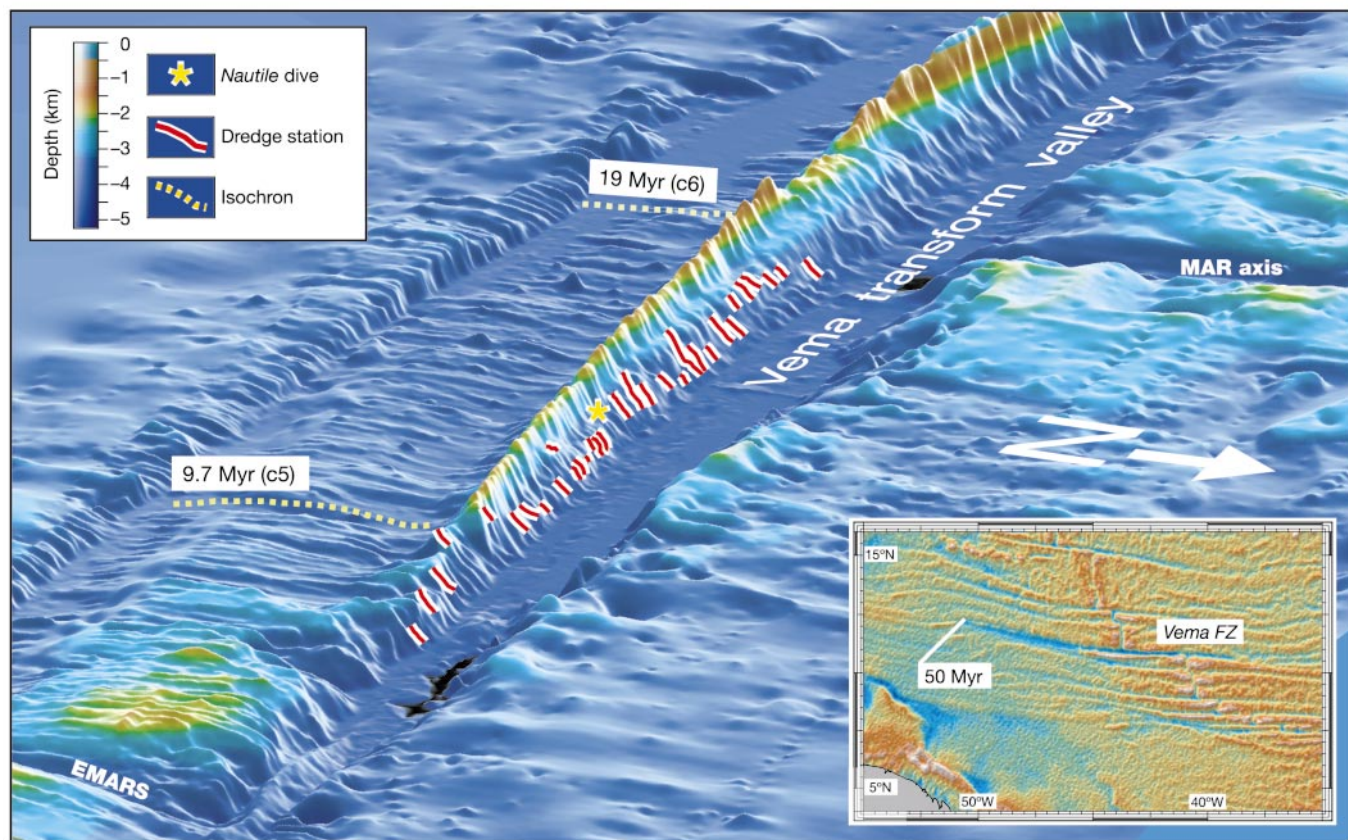


Figure 1 The Vema transform region viewed from the northeast. Figure is based on multibeam data obtained during cruises EW9305 (ref. 5), and S-19 and S-22 (RV *Akademik N. Strakhov*). The Vema lithospheric section (VLS) is clearly visible. Location of magnetic chrons C5 and C6, of the *Nautilie* submersible profiles, and of the rocks

sampled along the VLS, are indicated. The inset shows a satellite-derived gravity image of the Central Atlantic, where the Vema fracture zone (FZ), the transverse ridge and the long-lived EMARS can be identified.

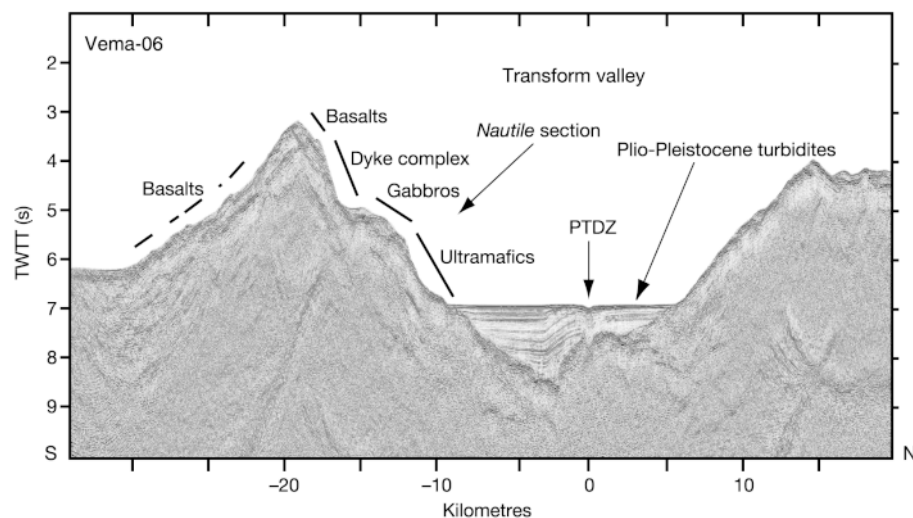


Figure 2 N-S multichannel seismic reflection profile (Vema-06) perpendicular to the Vema transform, above the *Nautilie* submersible sections. Profiles were obtained by the RV *Explora* using as sound source a tuned array of 28 air-guns (total volume of 60 litres, working pressure of 2,000 p.s.i.) towed 6 m below sea level, and as receiver a 3,150 m long streamer (120 channels spaced 25 m apart) towed 12 m deep. Lithology of the VLS is based on submersible observations and sampling. A reflector roughly parallel to the

southern slope of the Vema transverse ridge, about 0.5 s (two-way travel time, TWTT) below the sea floor, can be traced to the northern slope, where the boundary basalt/dyke complex was observed by submersible. This reflector probably marks the sub-bottom geometry of this boundary. PTDZ (Principal Transform Displacement Zone) points to the trace of strike-slip motion in the sediments filling the transform valley. North of this trace is the African plate; south of this trace is the South American plate.

elements by ion probe. The chromium number ($\text{Cr\#} = 100\text{Cr}/(\text{Al} + \text{Cr})$) of opx and spinel can be interpreted in terms of extent of melting undergone by these mantle rocks; this is because, during melting, Al partitions with the melt while Cr stays with the residue^{11–13}. We excluded spinels affected by secondary reactions with melts or fluids¹⁴. The Ti/Zr ratio of cpx can also be used to estimate degree of melting, because of the different partition coefficients of Ti and Zr during melting¹⁰. We calibrated our spinel and opx Cr# data in terms of degrees of melting assuming as a source a fertile, pyrolite-like peridotite^{13,15}. Degrees of melting calculated from Ti/Zr of cpx and from Cr# of opx and spinel correlate. They range along the VLS from 5.7% to 14.3%, with oscillations where the degree of melting appears to increase gradually and then decreases rapidly. These short-wavelength oscillations are superimposed on a long-term (>15 Myr) trend of increasing degree of melting with time, estimated on a regression line to range from $7 \pm 1.5\%$ about 19 Myr ago, to $13 \pm 2\%$ about 4 Myr ago (Fig. 3).

Given that the crust consists of frozen basaltic melt extracted from the mantle, and assuming no secondary crustal thinning (for instance, by core-complex tectonics), temporal variations of extent of melting should be reflected by variations of crustal thickness¹⁶, as discussed next.

Variations of gravity and crustal thickness

We compiled a free-air gravity map from a total of 31,872 shipboard gravity measurements, and combined this map with free-air gravity data obtained by satellite altimetry^{17,18}. The calculated mantle Bouguer anomaly (MBA) shows a prominent negative circular zone centred on the EMARS (Fig. 4c), consistent with the ‘bull’s eye’ MBA pattern detected elsewhere along the MAR², reflecting thicker crust at the segment’s centre. The MBA increases with age of the crust owing to cooling of the lithosphere. The thermal component of gravity, estimated assuming the present spreading rate (Fig. 4d), was subtracted from the MBA¹⁹; we thus obtained residual MBA (RMBA) images that reflect lateral variations of density and of crustal thickness (Fig. 4e). Given that lateral variations of crustal density are probably negligible, the RMBA reflects variations of crustal thickness (Fig. 4f). The RMBA pattern indicates a decrease of crustal thickness moving from the centre of the EMARS towards the Vema eastern ridge–transform intersection (RTI), and a minimal crustal thickness along the southern edge of the transform valley (Fig. 4), consistent with seismic refraction experiments^{20,21}. RMBA profiles along a flow line from the centre of the EMARS show (1) a long-range steady increase of crustal thickness from ~ 20 -Myr-old crust to the present (note that this pattern of steady crustal thickening would be enhanced if we were to take into account the

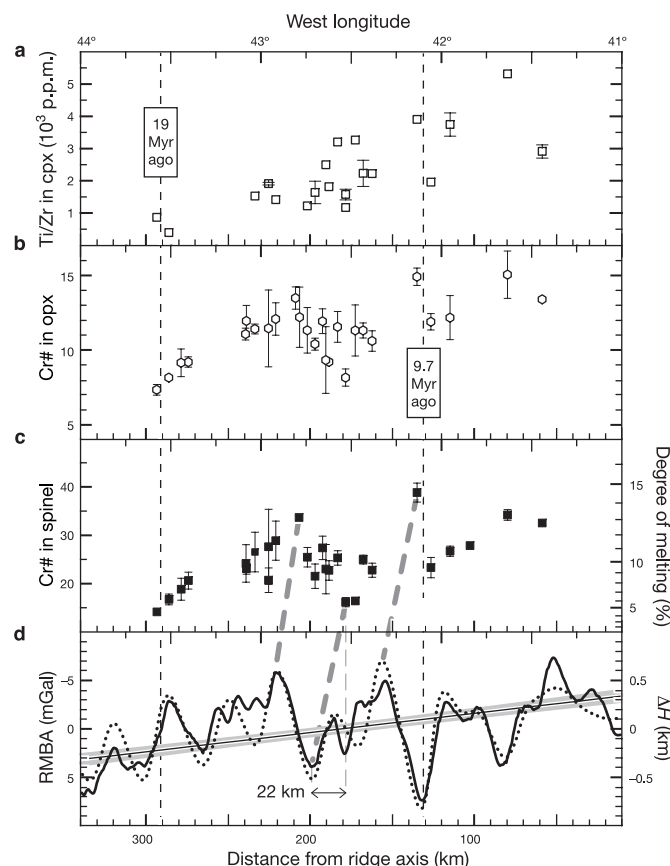


Figure 3 Temporal variations of mantle degree of melting and of crustal thickness. **a–c**, Lateral variations of Ti/Zr in cpx and Cr# in opx and spinel of mantle-derived peridotites sampled along the VLS. These curves, interpreted as reflecting variations of degree of melting, correlate ($R_{a,b} = 0.81$; $R_{b,c} = 0.97$). Major element mineral compositions were determined with a JEOL JXA 8600 and a CAMECA SX-100 electron probe; trace elements were determined with a CAMECA IMS 4F ion probe, using standard techniques. **d**, RMBA (thick solid line) and inferred variations of crustal thickness ΔH (dotted line) along the mid-point trace of EMARS shown in Fig. 4e. RMBA regression line (95% confidence interval) shows thickening of the crust moving towards the ridge axis.

The RMBA signal (scale inverted), indicating crustal thickness temporal variations, has been cross-correlated with the curve representing extent of melting of the peridotites. The maximum of the cross-correlation function (r) is reached when the RMBA curve is shifted eastward by 22 km ($r_0 = -0.1$, $r_{\max} = 0.7$). This is a good correlation, considering that geochemical and gravity data are from different locations along the ridge segment, that rock samples are not uniformly distributed, and that gravity was measured at sea level, with strong attenuation of high frequencies. Dashed lines in **c** and **d** show the suggested shift. Crustal ages based on ref. 7.

observed variations of spreading rate in correcting for the thermal component of gravity); (2) 3–4-Myr-long oscillations of crustal thickness, superimposed on the steady thickening. Each oscillation in the RMBA signal can be interpreted as reflecting a gradual increase of crustal thickness for ~3–4 Myr, followed by a rapid decrease (Fig. 3d).

The temporal variations of crustal thickness shown by the RMBA correlate with the variations in the mantle's degree of melting inferred from the peridotites, provided that the crustal thickness pattern is shifted eastward by ~22 km along the VLS (Fig. 3). This shift must reflect the different upwelling velocities of the melt and of the solid residue, allowing us to estimate the velocity of mantle advection below ridge axis.

Velocity of mantle upwelling

We estimated mantle advection velocity assuming rapid melt extraction and rise from the melting zone below the ridge axis, in line with uranium-series disequilibrium data suggesting melt upwelling velocities of metres per year^{22–24}. When the melt freezes near the sea floor, forming the crust, and starts moving laterally with

the plate, its parent mantle will be starting its slow ascent from the melting depth (Fig. 5). When this mantle parcel is incorporated in the lithosphere and starts moving laterally, the basalt it generated is further along the spreading path. The horizontal distance along the lithospheric section between the basalt and its parent mantle, translated as time, approximates the time it took for the residual mantle to travel from the melting depth to the lithosphere.

The ~22-km westward displacement of the curve representing temporal variations of crustal thickness relative to the curve representing extent of melting of the mantle peridotites (Fig. 3) is equivalent to ~1.3 Myr. This is approximately the time it took a parcel of mantle to rise from the melting depth to the lithosphere. Applying the equations of ref. 25, we estimate an average depth of melting below the EMARS of ~35 km; the mantle average upwelling rate from the melting depth is then ~24.8 mm yr⁻¹, that is, significantly higher than the 15.6 mm yr⁻¹ average half spreading rate. Based on mid-ocean-ridge basalt (MORB) uranium-series disequilibrium data, solid mantle upwelling rates of 25 mm yr⁻¹ were estimated in the Juan de Fuca ridge²⁴ similar to spreading half-rate (30 mm yr⁻¹).

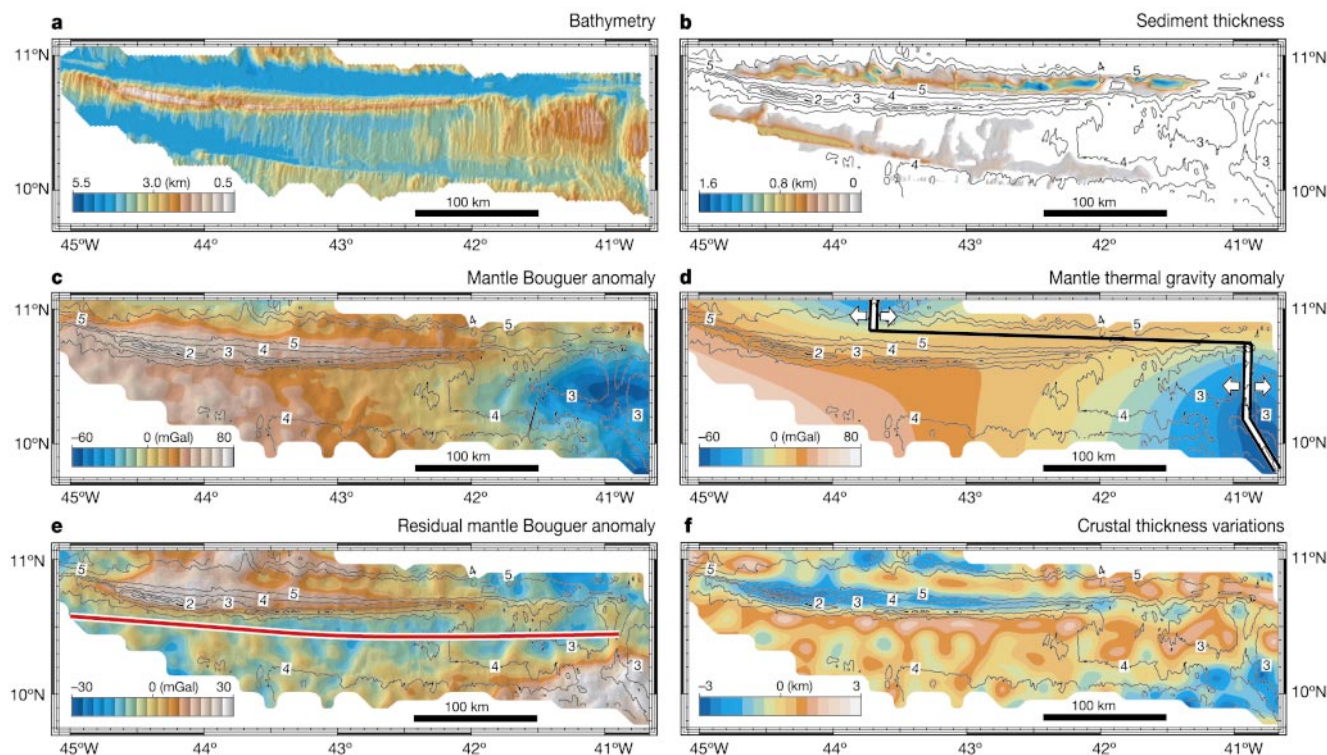


Figure 4 Gravity imagery in the Vema transform region. Shipboard gravity measurements, obtained during cruises EQUAP, C1702, V3206, K020A (ref. 48) and EW9305 (ref. 5), were corrected with the techniques outlined in ref. 49 leading to an accuracy of 2.6 mGal, with 95% confidence that anomalies >5 mGal are real and not noise. These data were combined with free-air gravity data obtained by satellite altimetry¹⁷ making use of input-output system theory¹⁸ in order to avoid edge effects in the computation of the MBA, due to lack of shipboard data outside the region of interest. **a**, Shaded relief image based on multibeam data. Mercator projection at 10° N. **b**, Sediment isopach map based on our multichannel seismic reflection profiles complemented by single-channel data described in ref. 48. Reflection times were converted to depths assuming uniform P-wave velocity of 2 km s⁻¹ throughout the sedimentary column. Bathymetric contours (interval, 1 km) mark the position of fracture zones and rift valleys. **c**, MBA map (zero level is arbitrary) generated by subtracting from the observed free-air anomalies the attraction of sea-floor topography and sediments, and of the crust–mantle interface, assuming 5-km-thick crust, and densities of 1,040, 1,900, 2,690 and 3,330 kg m⁻³ for sea water, sediments,

crust and mantle, respectively. **d**, Predicted gravity anomalies due to density variations caused by mantle thermal structure. Mantle temperatures were estimated by solving the steady-state advection–diffusion heat equation assuming 0 °C at the sea floor and 1,350 °C at 100 km of depth, using a three-dimensional domain of mantle flow calculations, with variable grid spacing (512 × 256 × 101), and highest grid resolution at the plate boundaries³². Density changes due to temperature were calculated assuming a thermal expansion coefficient of 3.25 × 10⁻⁵ per °C, chosen to match an averaged MBA profile, in a least-squares sense, along a 30-km-wide corridor centred on a flowline from EMARS midpoint. Mantle flow velocities were estimated assuming steady-state plate-thickening passive flow³³. **e**, Residual anomaly map. Thick red line marks the palaeo-midpoint trace of the EMARS. **f**, Inferred crustal thickness obtained after downward continuation of a low-pass filtered set (features with wavelength <25 km were filtered out) of the computed residual anomalies to a depth of 9 km below sea level, assuming a crust–mantle density contrast of 640 kg m⁻³.

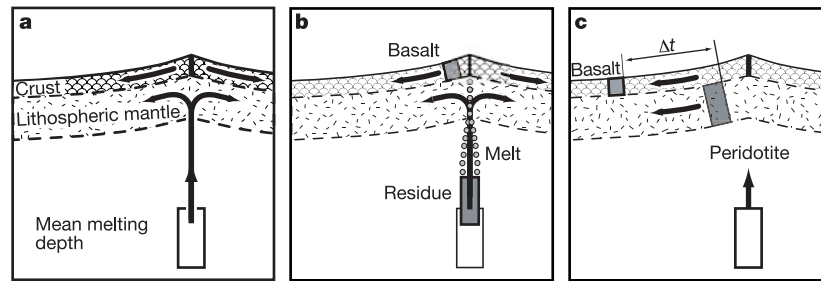


Figure 5 Differential upwelling of melt and of solid residual mantle beneath the ridge. Figure shows how a parcel of basaltic crust is displaced laterally relative to its parent mantle, and how mantle upwelling velocity can be estimated. **a**, Melting starts in the upwelling mantle within the melting interval. **b**, The melt rises rapidly and freezes, forming the basaltic crust, while the solid residual peridotitic mantle continues to rise. **c**, When the

residual peridotitic mantle joins the lithosphere and starts moving laterally with the plate, the basaltic crust it generated has moved further on with the plate. The distance between the basalt and its parent mantle, converted to time, approximates the time it took the peridotite to travel from the melting interval to the lithospheric plate.

However, peridotite geothermometry suggests that upwelling rates below the EMARS varied through time. Geothermometry calculations based on two pyroxene and single cpx equilibria in the VLS peridotites²⁶ showed equilibration temperatures that appear to correlate with the degree of melting. Equilibration temperatures are probably affected by the cooling rate of the system during ascent, which determines the extent to which coexisting phases can reach equilibrium at a given temperature; they are therefore related to the mantle's advection velocity, with higher values during faster upwelling²⁷. Therefore, the variations of equilibration temperatures we found along the VLS are probably related to variations of advection velocity below the ridge, with higher velocity associated with higher degree of melting.

Passive versus dynamic mantle upwelling and melting

We now seek to explain first, the short-wavelength (3–4-Myr)

oscillations of the mantle's degree of melting as it upwells below the EMARS, and of the thickness of the crust created at that segment; and second, the long-term (~20-Myr) steady increase of the mantle's degree of melting and of crustal thickness. Any interpretation must take into account the fact that the VLS (along which the variations of the mantle's extent of melting were estimated) formed originally towards the northern edge of the ~80-km-long EMARS, whereas variations of crustal thickness were estimated along a spreading flow line from the midpoint of the segment (Fig. 4e). Current models of slow-spreading mid-ocean ridges call for magma injection to be highest towards the centre of a segment^{28,29}, consistent with the 'bull's-eye' negative MBA (Fig. 4c) centred on the EMARS.

The fluctuations of magma supply inferred from the oscillations of crustal thickness can result from two processes: (1) periodic magma pulses at the surface ('magma solitons'), even when melt

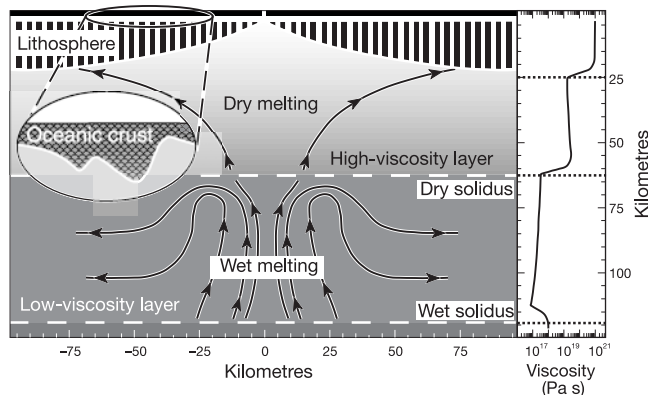


Figure 6 Model of mantle upwelling and crustal accretion consistent with the observed variations of extent of melting and crustal thickness at the EMARS. Arrows show direction of mantle flow. This crustal accretion pattern results from mantle upwelling induced by non-uniform sub-ridge mantle rheology, with a low-viscosity layer between two layers with higher viscosities⁴². The viscosity profile (right panel) describes the effects of dehydration, of melt in the matrix, and of change in creep mechanism. Water in the oceanic upper mantle deepens the peridotite solidus and the zone of partial melting relative to anhydrous mantle. Melt productivity during hydrous melting in the low-viscosity layer is low (1–2%); after water exhaustion melting continues above the 'dry solidus' in the high-viscosity layer, with higher melt production rates^{40,41}. High viscosity precludes compositional and thermal buoyant flow in the dry melting region, where solid flow is mostly driven by plate separation (corner flow). The low-viscosity layer favours intermittent plume-like enhanced upwelling and melting.

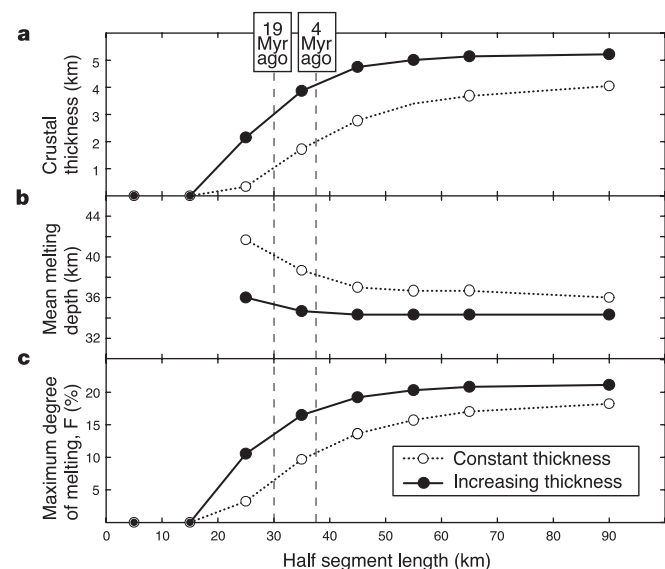


Figure 7 Extent of melting versus ridge half segment length when a ridge intersects a long-offset slow-slip transform such as Vema. **a**, Crustal thickness; **b**, mean melting depth; **c**, maximum degree of melting. Numerical experiments on melt production at ridge segment midpoint were based on ref. 50, using either constant or increasing lithospheric thickness passive-flow models. The length of the EMARS increased from 60 to 76 km during the time interval covered by our study, that is, from 19 to 4 Myr ago. The lithosphere thickening model appears to fit our data better.

production in the mantle is constant³⁰; and (2) fluctuations in extent of melting at the mantle source.

As we found variations not only of crustal thickness but also of the mantle's degree of melting (Fig. 3), the thermal structure and/or composition of the upwelling mantle must have varied through time. We infer a constant composition of the mantle source, based on the nearly constant Nd isotopic ratio and degree of enrichment shown by basaltic glasses sampled along the VLS (A.C. *et al.*, manuscript in preparation). Thus, both the short-wavelength oscillations and the long-term variation imply corresponding variations of mantle potential temperature and/or upwelling velocity.

We consider whether these results are compatible with a 'passive' model, where mantle upwelling and melting are caused solely by plate separation at ridge axis, or whether a dynamic component of mantle flow is required, driven by thermal and compositional buoyancy. Passive mantle flow is shaped by viscous drag from rigid plates that move apart. Predicted flow velocity is close to spreading half-rate, but is affected also by plate geometry close to the ridge axis^{31,32}. Where the plate thickens rapidly with distance from ridge axis, upwelling is narrowly focused and vertical velocity can exceed the half spreading rate³³.

Dynamic flow models require low viscosities in the melting region, and buoyancy forces due to thermal and compositional lowering of density caused by melting, that is, by decrease in the residue of Fe/Mg ratio and of dense phases such as garnet^{34–36}.

Short-wavelength oscillations

Oscillations of crustal thickness with periods of a few Myr have also been inferred from gravity elsewhere in the MAR^{37,38}; therefore, they may reflect general mechanisms of crustal formation at slow-spreading ridges. Numerical models for slow ridges that include non-uniform mantle rheology and thermal and compositional buoyancy predict that melting, and consequently crustal thickness, undergo oscillations, the frequency and amplitude of which depend on mantle viscosity^{36,39}.

The observed oscillations can be explained if we assume a sub-ridge low-viscosity zone ($\sim 10^{18}$ Pa s) underlying a higher-viscosity layer ($\sim 10^{20}$ Pa s) (Fig. 6). This rheological stratification may be due to extraction of H₂O from the upwelling mantle in the deeper part (>60 km) of the melting interval ('wet melting' interval). H₂O, present in significant quantities ($\sim 0.1\%$) in the oceanic upper mantle⁴⁰, partitions with the melt during 'wet melting', driving up the viscosity of the residue by two orders of magnitude^{40,41}. The high-viscosity residue continues to upwell and undergoes 'dry' melting above ~ 60 km (Fig. 6). High viscosity precludes dynamic mantle flow³⁶, so solid flow within the upper layer is mostly driven by plate separation (passive flow). However, the low-viscosity layer allows compositional and thermal buoyant flow^{36,42}. A possible model is that, as the mantle moves away from the ridge axis, it becomes denser owing to cooling and to the decreasing fraction of melt it contains because of the melt's upward migration, thus overcoming compositional buoyancy. According to numerical models, a dense plume forms and sinks periodically into the mantle, giving rise to intermittent convective rolls, periodic enhanced upwelling and increased temperature at the top of the low-viscosity layer, and, therefore, to increased melting below the ridge axis. The periodicity depends on the thickness of the low-viscosity layer⁴²; its assumed thickness (Fig. 6) is consistent with our observed 3–4-Myr oscillations.

Is the MAR becoming hotter?

The observed long-term trend of the mantle's increasing degree of melting and of the thickening of the crust suggests that the mantle upwelling beneath the MAR has become gradually hotter over the past 20 Myr, during which the spreading half rate slowed down significantly⁷ (Fig. 3). We offer two explanations. The first calls for variations of the length of the EMARS with time. The second, of

broader significance, implies long-term thermal pulses in the central Atlantic mantle. The first hypothesis is based on satellite gravity data (Fig. 1) showing a gradual lengthening of the EMARS since about 50 Myr ago owing to changes in ridge/transform geometry related to a westward migration of the eulerian pole between Africa and South America. RMBA and segment length are related, the longest segments having the largest RMBA^{2,43,44}. However, this correlation is interpreted as being due to stronger focusing of melt at segment midpoint in longer segments, while melt production appears to be independent of segment length⁴⁵. Segment length is important, however, because midpoints of longer segments are least affected by the cold-edge effect of a long-offset transform. Numerical experiments on the mantle's degree of melting versus segment length when a segment meets a long-offset, slow-slip transform such as Vema, do indeed show increasing degrees of melting below the segment's centre as the segment lengthens (Fig. 7).

We prefer, however, the second explanation—that is, steady increase of mantle temperature below the MAR, resulting in increasing melt production despite decreasing spreading rates. The idea of a gradually hotter northern MAR is supported by crustal thinning with age elsewhere in the MAR between 29° and 40°N (refs 38, 45). MAR topography and geochemistry suggest an overall decrease of sub-ridge mantle temperature from the Iceland/Azores swells to the equatorial 'cold' region. We speculate that the gradual heating of the MAR in the past 20 Myr may be related to hot sub-asthenospheric mantle flowing southward⁴⁶, consistent with seismic anisotropy in the North Atlantic⁴⁷.

Summary

Extending our results to slow-spreading ridges in general, we come to the following conclusions. (1) 3–4-Myr oscillations of degree of melting and crustal thickness are caused by intermittent buoyant 'active' convective upwelling, due to non-uniform mantle rheology beneath the ridge. Thus, the slow-spreading oceanic lithosphere is not only segmented along-axis; its structure also varies systematically normal to the ridge axis. (2) The solid mantle average upwelling velocity beneath the ridge can be significantly higher than the spreading half-rate, and can vary through time, being highest during intervals of high degree of melting. This supports the idea of an 'active' component in mantle advection and melting. (3) Much of the central MAR has become gradually hotter over the past 20 Myr, possibly owing to southward flow of hot mantle. □

Received 8 October 2002; accepted 28 March 2003; doi:10.1038/nature01594.

- Macdonald, K. C. *et al.* A new view of the mid ocean ridge from the behaviour of ridge axis discontinuities. *Nature* **335**, 217–225 (1988).
- Lin, J., Purdy, G. M., Shouten, H., Sempere, J. C. & Zervas, C. Evidence from gravity data for focused magmatic accretion along the Mid-Atlantic Ridge. *Nature* **344**, 627–632 (1990).
- Phipps Morgan, J. & Parmentier, E. M. Crenulated seafloor: Evidence for spreading-rate dependent structure of mantle upwelling and melting beneath a mid-oceanic spreading center. *Earth Planet. Sci. Lett.* **129**, 73–84 (1995).
- Auzende, J. M. *et al.* Direct observation of a section through slow-spreading oceanic crust. *Nature* **337**, 726–729 (1989).
- Kastens, K. *et al.* The Vema transverse ridge (Central Atlantic). *Mar. Geophys. Res.* **20**, 533–556 (1998).
- Gasparini, L. *et al.* Time constraints on the emplacement of an uplifted sliver of lithosphere at the Vema transverse ridge (Central Atlantic). *J. Conf. Abstr.* **4**, 757 (1999).
- Cande, S. C., LaBrecque, J. L. & Haxby, W. F. Plate kinematics of the south Atlantic, chron C34 to the present. *J. Geophys. Res.* **93**, 13479–13492 (1988).
- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. Mineralogic variability of the uppermost mantle along mid ocean ridges. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
- Michael, P. J. & Bonatti, E. Peridotite composition from the North Atlantic; regional and tectonic variations and implications for partial melting. *Earth Planet. Sci. Lett.* **73**, 91–104 (1985).
- Johnson, K. T. M., Dick, H. J. B. & Shimizu, N. Melting in the oceanic upper mantle: An ion microprobe study of diopsides in abyssal peridotites. *J. Geophys. Res.* **95**, 2661–2678 (1990).
- Jaques, A. L. & Green, D. H. Anhydrous melting of peridotite at 0–15 kb pressure and the genesis of tholeiitic basalts. *Contrib. Mineral. Petrol.* **73**, 287–310 (1980).
- Mysen, B. O. & Kushiro, I. Compositional variations of coexisting phases with degree of melting of peridotite in the upper mantle. *Am. Mineral.* **62**, 843–865 (1977).
- Hellebrand, E., Snow, J. E., Dick, H. J. B. & Hofmann, A. W. Coupled major and trace elements as indicators of the extent of melting in mid-ocean-ridge peridotites. *Nature* **410**, 677–681 (2001).
- Seyler, M. & Bonatti, E. Regional-scale melt-rock interaction in Iherzolitic mantle in the Romanche fracture zone (Atlantic Ocean). *Earth Planet. Sci. Lett.* **146**, 273–287 (1997).

15. Brunelli, D. *Variazioni temporali nei processi di formazione di litosfera lungo le dorsali medio-oceaniche* Thesis, Univ. Bologna (2001).
16. Forsyth, D. W. Crustal thickness and the average depth and degree of melting in fractional melting models of passive flow beneath mid-ocean ridges. *J. Geophys. Res.* **98**, 16073–16079 (1993).
17. Sandwell, D. T. & Smith, W. H. F. Marine gravity anomaly from Geosat and ERS-1 satellite altimetry. *J. Geophys. Res.* **102**, 10039–10054 (1997).
18. Li, J. & Sideris, M. G. Marine gravity and geoid determination by optimal combination of satellite altimetry and shipborne gravimetry data. *J. Geodesy* **71**, 209–216 (1997).
19. Kuo, B. Y. & Forsyth, D. W. Gravity anomalies of the ridge-transform system in the South Atlantic between 31° and 34.5° S: Upwelling centers and variations in crustal thickness. *Mar. Geophys. Res.* **10**, 205–232 (1988).
20. Louden, K. E., White, R. S., Potts, C. G. & Forsyth, D. W. Structure and seismotectonics of the Vema fracture zone, Atlantic Ocean. *J. Geol. Soc. Lond.* **143**, 795–805 (1986).
21. Potts, C. G., White, R. S. & Louden, K. E. Crustal structure of Atlantic fracture zones: II, The Vema fracture zone and transverse ridge. *Geophys. J. R. Astron. Soc.* **86**, 491–513 (1986).
22. Spiegelman, M. & Kenyon, P. M. The requirements for chemical disequilibrium during magma migration. *Earth Planet. Sci. Lett.* **109**, 611–620 (1992).
23. Spiegelman, M. & Elliott, T. Consequences of melt transport for uranium series disequilibrium in young lavas. *Earth Planet. Sci. Lett.* **118**, 1–20 (1993).
24. Lundstrom, C. C., Gill, J., Williams, Q. & Perfit, M. R. Mantle melting and basalt extraction by equilibrium porous flow. *Science* **270**, 1958–1961 (1995).
25. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 183–280 (Geophysical Monograph 71, American Geophysical Union, Washington D.C., 1992).
26. Taylor, W. R. An experimental test of some geothermometer and geobarometer formulations for upper mantle peridotites with application to the thermobarometry of fertile lherzolite and garnet websterite. *Neues Jb. Mineral. Abh.* **172**, 381–408 (1998).
27. Lasaga, A. C. in *Advances in Physical Geochemistry* Vol. 3 (ed. Saxena, S. K.) 81–114 (Springer, New York, 1983).
28. Magde, L. S., Sparks, D. W. & Detrick, R. S. The relationship between buoyant mantle flow, melt migration, and gravity bull's eyes at the Mid-Atlantic Ridge between 33°N and 35°N. *Earth Planet. Sci. Lett.* **148**, 59–67 (1997).
29. Magde, L. S., Barclay, A. H., Toomey, D. R., Detrick, R. S. & Collins, J. A. Crustal magma plumbing within a segment of the Mid Atlantic Ridge, 35°N. *Earth Planet. Sci. Lett.* **175**, 55–67 (2000).
30. Scott, D. R. & Stevenson, D. J. Magma solitons. *Geophys. Res. Lett.* **11**, 1161–1164 (1984).
31. Reid, I. & Jackson, H. R. Oceanic spreading rate and crustal thickness. *Mar. Geophys. Res.* **5**, 165–172 (1981).
32. Phipps Morgan, J. & Forsyth, D. W. Three-dimensional flow and temperature perturbations due to a transform offset: Effects on oceanic crustal and upper mantle structure. *J. Geophys. Res.* **93**, 2955–2966 (1988).
33. Blackman, D. K. & Forsyth, D. W. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. K.) 311–326 (Geophysical Monograph 71, American Geophysical Union, Washington D.C., 1992).
34. Buck, W. R. & Su, W. Focused mantle upwelling below mid-ocean ridges due to feedback between viscosity and melting. *Geophys. Res. Lett.* **16**, 641–644 (1989).
35. Scott, D. R. & Stevenson, D. J. A self-consistent model of melting, magma migration and buoyancy-driven circulation beneath mid-ocean ridges. *J. Geophys. Res.* **94**, 2973–2988 (1989).
36. Sotin, C. J. & Parmentier, E. M. Dynamical consequences of compositional and thermal density stratification beneath spreading centers. *Geophys. Res. Lett.* **16**, 835–838 (1989).
37. Pariso, J. E., Sempere, J. C. & Rommevaux, C. Temporal and spatial variations in crustal accretion along the Mid-Atlantic Ridge (29°–31°30' N) over the last 10 m.y.: Implications from a 3-D gravity study. *J. Geophys. Res.* **100**, 17781–17794 (1995).
38. Tucholke, B. E. *et al.* Segmentation and crustal structure of the western Mid-Atlantic Ridge flank, 25°25'–27°10' N and 0–29 m.y. *J. Geophys. Res.* **102**, 10203–10223 (1997).
39. Sparks, D. W., Parmentier, E. M. & Phipps Morgan, J. Three-dimensional convection beneath a segmented spreading center: Implications for along-axis variations in crustal thickness and gravity. *J. Geophys. Res.* **98**, 21977–21995 (1993).
40. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
41. Braun, M. G., Hirth, G. & Parmentier, E. M. The effect of deep damp melting on mantle flow and melt generation beneath mid-ocean ridges. *Earth Planet. Sci. Lett.* **176**, 339–356 (2000).
42. Choblet, G. & Parmentier, E. M. Mantle upwelling and melting beneath slow-spreading centers: Effects of variable rheology and melt productivity. *Earth Planet. Sci. Lett.* **184**, 589–604 (2001).
43. Neumann, G. A. & Forsyth, D. W. The paradox of the axial profile: Isostatic compensation along the axis of the Mid-Atlantic Ridge? *J. Geophys. Res.* **98**, 17891–17910 (1993).
44. Detrick, R. S., Needham, H. D. & Renard, V. Gravity anomalies and crustal thickness variations along the Mid-Atlantic Ridge between 33°N and 40°N. *J. Geophys. Res.* **100**, 3767–3787 (1995).
45. Hoof, E. E. E., Detrick, R. S., Toomey, D. R., Collins, J. A. & Lin, J. Crustal thickness and structure along three contrasting spreading segments of the Mid-Atlantic Ridge, 33.5°–35°N. *J. Geophys. Res.* **105**, 8205–8226 (2000).
46. Shen, Y. *et al.* Seismic evidence for a tilted mantle plume and north-south mantle flow beneath Iceland. *Earth Planet. Sci. Lett.* **197**, 261–272 (2002).
47. Yang, X. & Fischer, K. M. Constraints on North Atlantic upper mantle anisotropy from S and SS phases. *Geophys. Res. Lett.* **21**, 309–312 (1994).
48. Prince, R. A. & Forsyth, D. W. Horizontal extent of anomalously thin crust near the Vema fracture zone from the three-dimensional analysis of gravity anomalies. *J. Geophys. Res.* **93**, 8051–8063 (1988).
49. Huang, M., Guan, Z., Zhai, G. & Ouyang, Y. On the compensation of systematic errors in marine gravity measurements. *Mar. Geodesy* **22**, 183–194 (1999).
50. Bonatti, E. *et al.* Steady-state creation of crust-free lithosphere at cold spots in mid-ocean ridges. *Geology* **29**, 979–982 (2001).

Acknowledgements We thank A. Peyve and co-workers at the Geology Institute, Russian Academy of Science, and the captain and crew of the RV *Akademic N. Strakhov* for help with field work. We thank K. Kastens for providing cruise EW9305 gravity data; W. R. Buck and J. Karson for comments on the manuscript; and D.W. Forsyth for providing programs useful in processing gravity data. This work was supported by the Consiglio Nazionale Ricerche and the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.B. (enrico.bonatti@bo.ismar.cnr.it).

Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease

Hironori Ueda¹, Joanna M. M. Howson^{1*}, Laura Esposito^{1*}, Joanne Heward^{2*}, Hywel Snook¹, Giselle Chamberlain¹, Daniel B. Rainbow¹, Kara M. D. Hunter¹, Annabel N. Smith¹, Gianfranco Di Genova^{1†}, Mathias H. Herr^{1†}, Ingrid Dahlman^{1†}, Felicity Payne¹, Deborah Smyth¹, Christopher Lowe¹, Rebecca C. J. Twells¹, Sarah Howlett¹, Barry Healy¹, Sarah Nutland¹, Helen E. Rance¹, Vin Everett¹, Luc J. Smink¹, Alex C. Lam¹, Heather J. Cordell¹, Neil M. Walker¹, Cristina Bordin^{1†}, John Hulme¹, Costantino Motzo³, Francesco Cucca³, J. Fred Hess⁴, Michael L. Metzker^{4†}, Jane Rogers⁵, Simon Gregory⁵, Amit Allahabadia^{2†}, Ratnasingam Nithiyananthan², Eva Tuomilehto-Wolf⁶, Jaakko Tuomilehto^{6,7}, Polly Bingley⁸, Kathleen M. Gillespie⁸, Dag E. Undlien^{9†}, Kjersti S. Rønningen¹⁰, Cristian Guja¹¹, Constantin Ionescu-Tîrgoviște¹¹, David A. Savage¹², A. Peter Maxwell¹³, Dennis J. Carson¹⁴, Chris C. Patterson¹⁵, Jayne A. Franklyn², David G. Clayton¹, Laurence B. Peterson¹⁶, Linda S. Wicker¹, John A. Todd¹ & Stephen C. L. Gough²

*These authors contributed equally to this work

Genes and mechanisms involved in common complex diseases, such as the autoimmune disorders that affect approximately 5% of the population, remain obscure. Here we identify polymorphisms of the cytotoxic T lymphocyte antigen 4 gene (*CTLA4*)—which encodes a vital negative regulatory molecule of the immune system—as candidates for primary determinants of risk of the common autoimmune disorders Graves' disease, autoimmune hypothyroidism and type 1 diabetes. In humans, disease susceptibility was mapped to a non-coding 6.1 kb 3' region of *CTLA4*, the common allelic variation of which was correlated with lower messenger RNA levels of the soluble alternative splice form of *CTLA4*. In the mouse model of type 1 diabetes, susceptibility was also associated with variation in *CTLA-4* gene splicing with reduced production of a splice form encoding a molecule lacking the CD80/CD86 ligand-binding domain. Genetic mapping of variants conferring a small disease risk can identify pathways in complex disorders, as exemplified by our discovery of inherited, quantitative alterations of *CTLA4* contributing to autoimmune tissue destruction.

An understanding of the mechanisms underlying common diseases, which are caused by the combined effects of many genetic and environmental factors, will contribute to the development of future preventative and disease-modifying therapies. DNA variants of the human genome, which alter physiological pathways involved in susceptibility to multifactorial disease, are by definition primary causal risk factors. Progress, however, has been limited. Relatively few disease gene variants have been identified¹, probably due to the small risks associated with individual variants and an underestimation of how large studies have to be to obtain reliable results^{1,2}. The common autoimmune diseases are presumed to result from a failure in the mechanisms of the immune system that establish and maintain non-responsiveness or tolerance to self³. Our study implicates common allelic variation in the expression levels of alternative splice forms of a key negative regulator of the T lymphocyte immune response, *CTLA-4* (refs 4, 5), as a primary determinant of susceptibility to autoimmune disease.

Graves' disease affects 0.5% of western populations, and results

from the presence of autoantibodies to the thyrotropin receptor, leading to over-activity of the thyroid gland. Autoimmune hypothyroidism (AIH; 1% prevalence) is caused by a T-lymphocyte-mediated inflammatory destruction of thyrocytes within the thyroid gland, leading to thyroid hormone deficiency. Type 1 diabetes (T1D; 0.4% prevalence) is caused by insulin deficiency, associated with a T-cell inflammation of the insulin-producing cells of the pancreas.

Because Graves' disease, T1D, AIH and other autoimmune diseases such as rheumatoid arthritis commonly cluster in the same families, it is likely that they share some genetic background. Indeed, risk of Graves' disease, AIH and T1D has been associated with the same region on chromosome 2q33, which contains the T-lymphocyte regulatory genes *CD28*, *CTLA4* and inducible co-stimulator (*ICOS*)^{6–9}. Moreover, T1D susceptibility in the spontaneous mouse model of the disease, the nonobese diabetic (NOD) strain, has been mapped to the equivalent region of the mouse genome, on chromosome 1, containing the orthologues of

¹ Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 2XY, UK; ² Division of Medical Sciences, University of Birmingham, Birmingham, B15 2TT, UK; ³ Dipartimento di Scienze Biomediche e Biotecnologie, Università di Cagliari, Cagliari, Italy; ⁴ Merck Research Laboratories, West Point, Pennsylvania 19486, USA; ⁵ Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK; ⁶ Diabetes and Genetic Epidemiology Unit, National Public Health Institute, University of Helsinki, Helsinki, Finland; ⁷ Department of Public Health, University of Helsinki, Helsinki, Finland; ⁸ Diabetes and Metabolism Unit, Division of Medicine, University of Bristol, Bristol, BS10 5NG, UK; ⁹ Institute of Immunology, Rikshospitalet University Hospital and Institute of Medical Genetics, Ullevål University Hospital, University of Oslo, Oslo, Norway; ¹⁰ Laboratory of Molecular Epidemiology, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; ¹¹ Clinic of Diabetes, Institute of Diabetes, Nutrition and Metabolic Diseases 'N. Paulescu', Bucharest, Romania; ¹² Department of Medical Genetics, Queen's University Belfast, Belfast City Hospital, Belfast, BT9 7AB, UK; ¹³ Regional Nephrology Unit, Belfast City Hospital, Belfast, BT9 7AB,

UK; ¹⁴ Department of Child Health, Queen's University Belfast, Belfast, BT12 6BJ, UK; ¹⁵ Department of Epidemiology and Public Health, Queen's University Belfast, Belfast, BT12 6BJ, UK; ¹⁶ Department of Pharmacology, Merck Research Laboratories, Rahway, New Jersey 07065, USA.

† Present addresses: HIT Group, Tenovus Research Laboratories, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK (G.D.G.); Department of Paediatrics, Otto-Heubner-Centrum, Charité, Humboldt-University of Berlin, Germany (M.H.H.); Karolinska Institute, Department of Medicine, Huddinge University Hospital, Stockholm, Sweden (I.D.); Hutchison/MRC Research Centre, Hills Road, Cambridge, CB2 2XZ, UK (C.B.); Department of Molecular & Human Genetics, Baylor Genome Sequencing Center, Baylor College of Medicine, N1409, One Baylor Plaza, Houston, Texas 77030, USA (M.L.M.); Division of Clinical Sciences, University of Sheffield, Northern General Hospital, Sheffield, S5 7AU, UK (A.A.); Institute of Medical Genetics, University of Oslo, N-0315 Oslo, Norway (D.E.U.).

Table 1 Association of *CTLA4* SNPs in 3,671 type 1 diabetic families

SNP/allele	RR	95% confidence intervals	P value	Transmission to affected offspring (%)	P_{TDT}
CTAF343/T	1.18	1.08–1.27	0.00009	54.0	0.00009
rs1863800/C	1.16	1.09–1.25	0.00001	53.8	0.00001
MH30/G	1.16	1.09–1.24	8.0×10^{-6}	53.7	8.3×10^{-6}
+49/G	1.09	1.02–1.16	0.013	52.1	0.013
CT60/G	1.14	1.07–1.21	0.00006	53.3	0.00006
JO31/G	1.12	1.05–1.20	0.00090	52.9	0.00091
JO30/G	1.18	1.10–1.27	5.7×10^{-6}	54.2	5.8×10^{-6}
JO27_1/T	1.17	1.09–1.25	0.00003	53.8	0.00003
CTIC154_1/C	1.01	0.89–1.14	0.91	50.2	0.91

Relative risk (RR) was calculated from the coefficients of the regression equation⁴⁵, and P values are for the null hypothesis of no association of the marker. For comparison, the P_{TDT} (where TDT is the transmission/disequilibrium test) probability values⁴⁶ are provided along with percentage transmission from heterozygous parents to affected offspring.

CD28, *CTLA4* and *ICOS*^{10–12}. However, the causal variants in neither humans nor mice have been localized or identified. All three genes (*CD28*, *CTLA4* and *ICOS*) are strong candidates for the source of the disease risk associated with this chromosome region and its mouse orthologue, because they have critical roles in T-cell activation and in autoimmune disease^{4,13–16}.

Human genetics

Initially, the sequence of most of the 300-kilobase (kb) *CD28–CTLA4–ICOS* region was unknown. Hence, the possibility of the presence of other candidate genes could not be excluded, and the polymorphism content was also unknown. We sequenced 330 kb of human DNA containing *CD28*, *CTLA4* and *ICOS*, and where possible re-sequenced segments from 32 individuals to identify the polymorphisms (Supplementary Information A). We also sequenced the orthologous mouse *Ctla4–Icos* region to help establish the gene content and to aid our search for the mouse T1D gene (*Id5.1*; refs 10–12) that exists in this region of mouse chromosome 1 (see below).

CD28, *CTLA4* and *ICOS* were the only functional genes in the region (Supplementary Information A). No polymorphisms that altered amino acids were found in any of the exons of the three genes, except for the known +49G>A single nucleotide polymorphism (SNP) (Thr17Ala) in exon 1 of *CTLA4* (Supplementary Information A). A total of 108 SNPs and the *CTLA4* (AT)_n-3' untranslated region (UTR) marker were genotyped in 384 cases of Graves' disease and 652 controls to determine their allele frequen-

cies (Fig. 1; see also Supplementary Information A and B). If a marker allele is more or less frequent in cases compared to controls it can be categorized as being associated with disease. This is due to association, or linkage disequilibrium (LD), of the marker allele with an allele of the causal variant, or to the marker itself being a causal variant. Significant association was observed in the region (Fig. 1). *CTLA4* and the 5' part of *ICOS* were contained within a 100-kb subregion, or block¹⁷, of strong intermarker LD flanked by two localized regions of low LD, one between *CD28* and *CTLA4*, the other in the first intron of *ICOS* (Fig. 1; see also Supplementary Fig. B1). The causal variant was most probably located within the 100-kb region of LD containing *CTLA4* and the 5' region of *ICOS*, with the association reaching a maximum in a region immediately 3' of the *CTLA4* structural gene ($P = 1.6 \times 10^{-6}$; odds ratio of 1.56). Outside this LD block, disease association declined rapidly. None of the five SNPs tested in *CD28* or the 14 SNPs in *ICOS* were associated with disease (Fig. 1; see also Supplementary Figs B2–B4 and Table B7).

Within the *CTLA4* LD block the pattern of disease association had three main peaks (from left to right in Fig. 1) with 23 out of 78 (29.5%) SNPs showing no association ($P > 0.05$). The first peak ($P \approx 10^{-5}$) was 20–35 kb 5' of the *CTLA4* ATG codon (including the marker MH30). The second ($P \approx 10^{-6}$) was 0.2–6.3 kb 3' of the previously reported end of the *CTLA4* transcript and included SNPs CT60, JO31, JO30 and JO27_1. The third ($P \approx 10^{-4}$) was 49–55 kb 3' of the *CTLA4* transcript end and 8–13 kb 5' of the *ICOS* ATG codon, and comprised seven markers, including CTBC217_1. Regression modelling¹⁸ indicated that the third peak was an unlikely location for the causal variant because once the most-associated marker from the middle peak (CT60) was included in the model, the most-associated marker from the third did not improve the model (Supplementary Figs B2–B4). Furthermore, none of the previously known polymorphisms, -319C>T (CT44), +49G>A, +1,822T>C (CT55)⁷ and (AT)_n-3' UTR, could explain the association of the region with Graves' disease (Supplementary Table B2).

To distinguish between the remaining two association peaks, the seven SNPs associated most closely with Graves' disease along with the +49G>A SNP were genotyped in a larger data set (a further 288 Graves' disease cases and 192 controls; Supplementary Table B3). Regression analyses showed that the first peak was less likely to harbour the causal variant than the second peak (Supplementary Information B). The most-associated marker remained CT60 (odds ratio of 1.51, 95% confidence interval of 1.31–1.75; $P = 2.72 \times 10^{-8}$) but it could not be distinguished in the regression analyses from its neighbouring SNPs: JO31, JO30 and JO27_1. CT60 is a common SNP with 63.4% of 1,316 Graves' disease patient chromosomes and 53.2% of 1,646 control chromosomes having the susceptible G allele. Compared with the protective CT60 A/A genotype, the A/G and G/G genotypes had odds ratios of 1.59 (1.19–2.13) and 2.32 (1.71–3.15), respectively. In controls the A/A, A/G and G/G genotypes had frequencies of 22.8%, 48.0% and 29.2%, and in cases, 13.7%, 45.7% and 40.6%, respectively. Conversely, relative to the disease-predisposing G/G genotype, A/G and

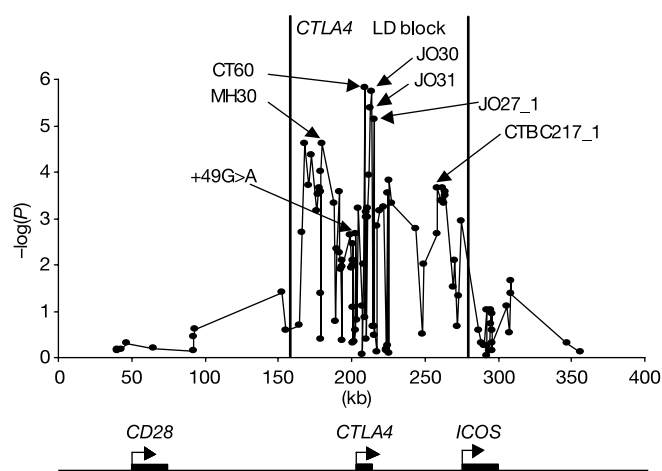


Figure 1 Association of the *CD28–CTLA4–ICOS* region with Graves' disease. The log P values (y-axis) of the differences in allele frequencies of 108 SNPs and the *CTLA4* (AT)_n-3' UTR microsatellite marker between cases of Graves' disease ($n = 384$) and controls ($n = 652$) were plotted against physical distance (x-axis). Vertical lines indicate the two boundaries of the LD blocks as shown in Supplementary Fig. B1. Lines are drawn between points only for the sake of clarity, as the disease association between markers is not linear.

A/A had odds ratios of 0.68 (0.54–0.86) and 0.43 (0.32–0.59), respectively (Supplementary Information B). Regression analysis of haplotypes provided no evidence for haplotype-specific effects; there are two very common CTLA4 haplotypes, one susceptible for Graves' disease, the other, protective (Supplementary Table B4). This result along with the single-point-association analysis (Fig. 1 and Supplementary Table B1), a Graves' disease family-based study (Supplementary Table B1) and the two-locus regression results (Supplementary Figs B2–B4) suggested that a single common variant, one of the SNPs CT60, JO31, JO30 or JO27_1 in the 6.1-kb region 3' of *CTLA4*, is responsible for the disease association of the region.

The CT60 SNP was also associated with AIH (228 cases and 844 controls) to the same degree as Graves' disease (odds ratio of 1.45 (1.17–1.80); $P = 0.0005$). However, the effect was much weaker in T1D. In 3,671 T1D families we analysed nine markers: CTAF343, rs1863800 and MH30 in the first Graves' disease peak; CT60, JO31, JO30 and JO27_1 in the second peak, along with +49G>A, and CTIC154_1, a SNP in the 3' UTR of *ICOS*. Alleles that were positively disease-associated in Graves' disease were also associated with T1D, with the most-associated markers being MH30, JO30 and CT60; however, the effect was small (odds ratios ≈ 1.15 ; Table 1). Regression analysis failed to exclude the 5' *CTLA4* markers MH30 and rs1863800. However, +49G>A was rejected as the causal SNP, as were CTAF343, JO31 and the *ICOS* SNP CTIC154_1 (Supplementary Information B). Increased transmission of *CTLA4* alleles to siblings affected with T1D (53–54%, Table 1) was mirrored by under-transmission to unaffected siblings: for example, the G allele of CT60 was under-transmitted to unaffected offspring at 47.1% (989:1,110; $P = 0.009$), ruling out transmission ratio distortion and genotyping error as explanations for the association in T1D. The T1D data do raise the possibility that additional, more rare causal variants may exist in the region (Supplementary Table B5). Furthermore, neither the T1D nor Graves' disease data exclude the possibility that the common causal variant lies outside the 3' 6.1-kb region; for example, 5' of *CTLA4*. Nevertheless, taken together, the data suggest the presence of a common Graves' disease, T1D (designated previously *IDDM12*) and AIH locus in the 6.1-kb 3' region of *CTLA4*.

CTLA-4 gene expression and genotype

As the most disease-associated SNPs did not affect the CTLA-4 amino acid sequence, and the 6.1-kb region contained most of the CTLA-4 3' UTR (Supplementary Fig. A4), we examined the steady-state mRNA levels of the two known isoforms of CTLA-4: a full-length isoform (fICTLA-4) encoded by exon 1 (leader peptide), exon 2 (ligand-binding domain), exon 3 (transmembrane domain) and exon 4 (cytoplasmic tail), and a soluble form (sICTLA-4)^{19–21}, which lacks exon 3 (Fig. 2a). We used a combination of polymerase chain reaction (PCR) primers specific for the fICTLA-4 and sICTLA-4 mRNAs and allele-specific transcript quantification (ASTQ)²² to measure the relative amounts of the sICTLA-4 and fICTLA-4 mRNAs transcribed from the susceptible and protective haplotypes. Unstimulated CD4 T cells from four individuals were examined: three doubly heterozygous +49A/G-CT60 A/G individuals, all having one dose each of the protective (+49A-CT60A) and susceptible (+49G-CT60G) haplotype across the 6.1-kb region, and a +49G/A heterozygous/CT60 G/G homozygous individual, who carried a haplotype with a recombination event between +49 and CT60, corresponding to two doses of the susceptible 6.1-kb region haplotype. The sICTLA-4 mRNA level from the disease-protective haplotype (A allele at +49) was higher than that of the sICTLA-4 level transcribed from the susceptible haplotype in the samples from all three +49A/G-CT60 A/G heterozygous individuals, giving ratios of 2.0:1, 2.1:1 and 2.5:1 (Fig. 2b). These ratios were significantly different to the ratios of fICTLA-4 transcript production according to haplotype (1.3:1, 1.7:1, 1.4:1 and 1.5:1; $P = 0.007$; note that these

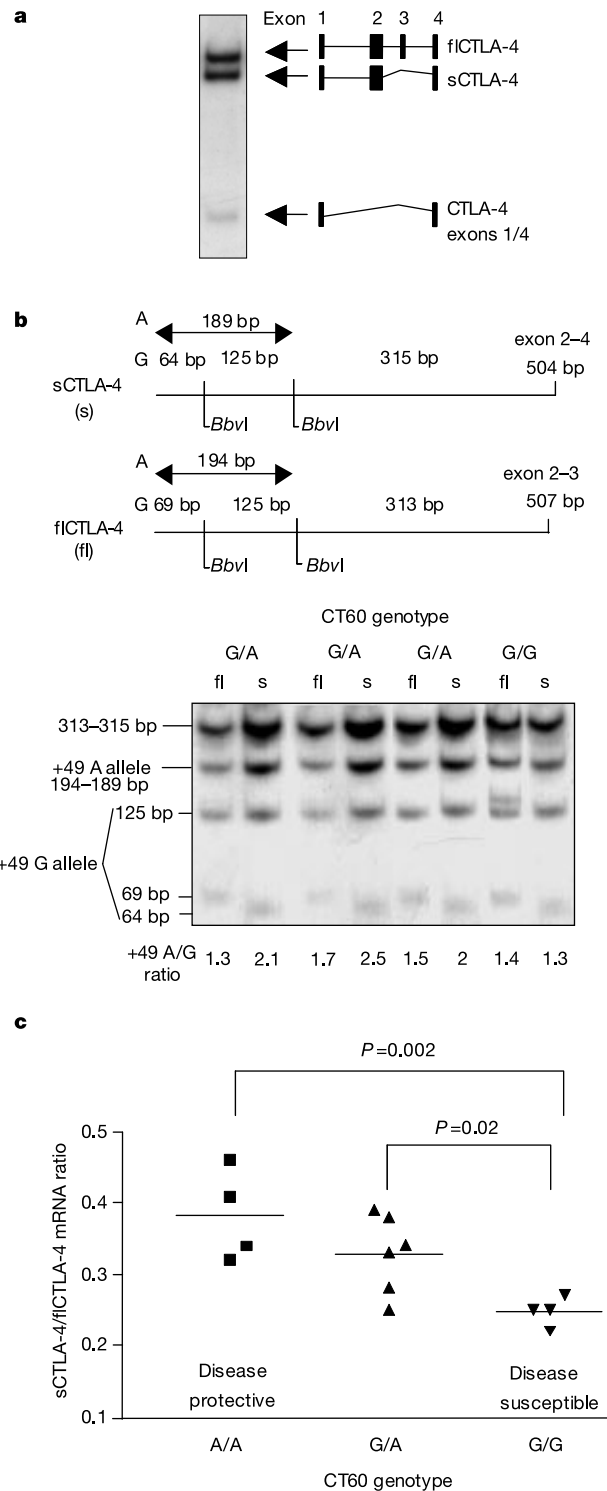


Figure 2 Expression of the human CTLA-4 mRNA isoforms correlates with genotype. **a**, Three alternative splice variants of CTLA-4 (fICTLA-4 (672 bp), sICTLA-4 (562 bp) and exon 1–4 (214 bp)) are detected using RT–PCR in PBMCs. **b**, ASTQ of fICTLA-4 and sICTLA-4 using the +49G/A restriction fragment length polymorphism in RNA(cDNA) derived from CD4 T cells of four healthy volunteers (three CT60 G/A, one CT60 G/G). Relative contributions of the +49 A and G allele haplotypes of fICTLA-4 and sICTLA-4 transcripts are expressed as a +49 A (189 or 194 bp) to G (125 bp) ratio. The smallest sized bands from the G allele (64 and 69 bp) were not included when calculating the ratio owing to their diffuse nature; thus, all +49 A/G ratios are overestimates such that a +49 A/G ratio of 1:1 is observed as a 1.5:1 ratio. **c**, Expression of sICTLA-4 relative to fICTLA-4 measured by real-time quantitative PCR of RNA purified from CD4 T cells of 14 healthy volunteers (four A/A, six A/G and four G/G).

ratios are equivalent to a ratio of 1:1 owing to the size difference in the ASTQ PCR bands; Fig. 2b). As would be expected if allelic variation of the 6.1-kb region was the cause of the difference in splicing efficiency, the individual who was homozygous-susceptible for the region showed a sCTLA-4 ratio of 1.3.

To extend these findings to individuals with all three genotypes at CT60, we developed a real-time PCR assay that specifically detected the mRNA abundance of the sCTLA-4 isoform. To account for the fact that variable numbers of CTLA-4-positive cells would be present in the peripheral blood of different individuals, flCTLA-4 was used to normalize the sCTLA-4 values. We found that the ratio of sCTLA-4 to flCTLA-4 mRNA splice forms in unstimulated CD4 T cells was 50% lower in CT60 G/G-positive disease-susceptible

individuals compared with A/A protected individuals (Fig. 2c; $P = 0.002$). Consistent with the semi-dominant mode of inheritance of the disease locus at *CTLA4*, the heterozygous genotype expressed more sCTLA-4 mRNA than the homozygous susceptible genotype ($P = 0.02$). Our results suggest that the 6.1-kb region determines the efficiency of the splicing and production of sCTLA-4, with the CT60G disease-susceptibility haplotype producing less sCTLA-4 transcript than the resistant CT60A haplotype.

Mouse CTLA-4

Recent genetic mapping has defined the location of the *Idd5.1* T1D disease locus to a 1.1 cM (2.1 megabase) region of chromosome 1 containing the human chromosome 2q33 orthologues *Ctla4*, *Icos*, *Als2cr19* and part of *Nrp2* (ref. 11) (L.S.W., manuscript in preparation). *Cd28* was excluded as the *Idd5.1* causal locus. In the *Idd5.1* region the C57BL/10 (B10) haplotype is protective for T1D whereas the NOD haplotype confers susceptibility to disease. Comparison of genomic sequence between the NOD and B10 strains of the two functional candidate genes *Ctla4* and *Icos* revealed only two exonic variants: one translationally silent change in exon 2 of *Ctla4*, and one coding variant causing a non-conservative amino acid change in the putative leader sequence encoded in exon 1 of *Icos*. No variants were found in either gene at any of the splice sites or branch points. Nevertheless, splicing of *Ctla4* was investigated in detail (Fig. 3).

We discovered a new splice form of mouse CTLA-4 with a fully intact open reading frame that encodes a transmembrane isoform of CTLA-4 with the CD80/CD86-binding domain missing, referred to here as the ligand-independent isoform (liCTLA-4; Fig. 3a–c). We also observed the transcripts for the previously described full-length and soluble isoforms (Fig. 3a–c). The liCTLA-4 transcript is formed by skipping exon 2, and has a fourfold increase in expression in both activated and resting spleen cells from the diabetes-resistant strain as compared with the NOD strain (Fig. 3d).

A literature survey and DNA sequence comparison indicated that the mouse CTLA-4 exon 2 SNP is the probable genetic cause of the variation in liCTLA-4 mRNA levels (Fig. 3e). In the human CD45 gene, substitution of the wild-type C allele with a mutant G allele by a SNP at base 77 of exon 4 reduces the function of an exonic splicing silencer (ESS) embedded in the exon 4 DNA sequence, resulting in alteration of the relative levels of functionally distinct isoforms of this important T-cell differentiation molecule²³. Comparison of the human CD45 gene exon 4 sequence and CTLA-4 gene exon 2 sequence, and their mouse and rat orthologues, identified the conserved wild-type sequence CNNNTGNATNNNCACCA/CNCANNANCNNNTGA (see Fig. 3e). In marked similarity to the human CD45 exon 4 SNP, the mouse *Ctla4* exon 2 SNP has an A at position 77 in the diabetes-resistant strains B10 and B6, whereas the NOD strain carries the mutant G allele. The G allele would lead to an increase in the inclusion of exon 2, forming flCTLA-4 mRNA, and therefore a reduction of the exon-2-negative transcript that encodes liCTLA-4. Eight of the conserved 18 nucleotides in this 31-base pair (bp) region differed in human *CTLA4*, which probably accounts for our inability to detect the liCTLA-4 transcript in human T cells (Fig. 2a). The rat *Ctla4* exon 2 sequence differed only by two nucleotides among the conserved 18 compared to the mouse sequence, at positions 75 and 77: both are T nucleotides versus C and A or C in ESS wild-type sequences (Fig. 3e). Rat spleen does not express detectable levels of the liCTLA-4 isoform (data not shown), supporting a critical role for position 77 for the splicing efficiency of liCTLA-4. Mutational studies of an ESS in the late pre-mRNA of bovine papillomavirus type 1 and also sequence comparisons with the few other known ESSs have demonstrated the importance of this central CCCCC sequence²⁴.

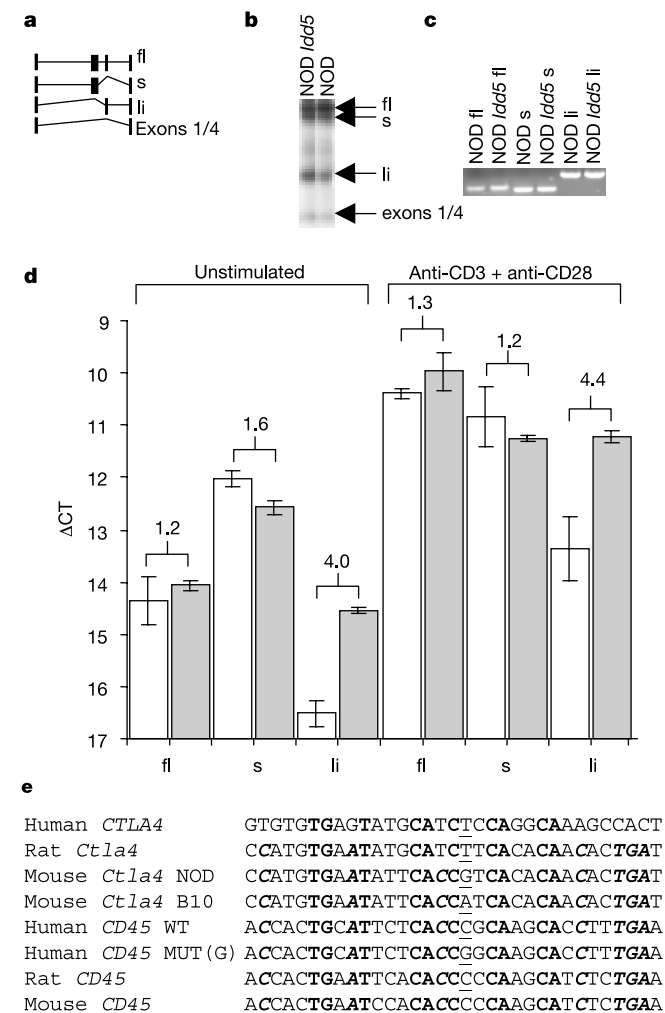


Figure 3 Sequence, splicing and expression of *Ctla4* in NOD mice and congenic strains. **a**, Schematic of the four splice forms of *Ctla4*. **b**, Autoradiogram of a Southern blot illustrating the splicing of CTLA-4 mRNA. Primers are in exons 1 and 4 of *Ctla4*, producing bands of 624, 514, 276 and 166 bp corresponding to flCTLA-4, sCTLA-4, liCTLA-4 and the exons 1/4 isoform, respectively. **c**, Agarose gel of the products generated by RT-PCR using primers specific to each CTLA-4 isoform mRNA. **d**, Expression of flCTLA-4, sCTLA-4 and liCTLA-4 mRNA by real-time quantitative PCR from purified splenic T cells before and 6 h after anti-CD3 (5 µg per mouse) plus anti-CD28 (2 µg per mouse) antibody stimulation *in vivo*. Open bars, NOD T cells; filled bars, diabetes-resistant NOD *Id5* T cells. Error bars represent 1 s.d. **e**, Sequence alignment of *CD45* and *CTLA4* flanking the SNP at base 77 of exon 4 of *CD45* and exon 2 of *CTLA4*. Residues in bold are conserved between all sequences, and residues in bold and italicized are conserved between all except those in which splicing is not observed (human and rat *CTLA4*). Position 77 is underlined.

Discussion

Our results highlight the power of positional genetics in the identification of key physiological DNA variants, and they provide evidence that certain common autoimmune diseases are caused in part by inherited changes in CTLA-4 expression that presumably increase T-cell self-reactivity. The importance of CTLA-4 in the counter-regulation of CD28 T-cell antigen receptor (TCR) activation of T cells is well established^{4,13–16}; however, the molecular basis of the interactions between CTLA-4 and CD28 is uncertain. CTLA-4, by means of its ligand-binding domain, may compete with CD28 for the receptors (CD80 and CD86) that they share. CTLA-4 might also deliver a negative signal to each T cell by inhibiting the TCR signalling complex, an activity that would not necessarily depend on ligand binding¹⁶. Furthermore, and non-exclusively, CTLA-4 is expressed constitutively by T cells that suppress the effector T-cell response; referred to as T-regulatory cells (T_R)^{4,25,26}. Depletion of T_R cells causes autoimmune disease, and the function of some T_R cells is dependent on CTLA-4^{27–29}. T_R cells may exert part of their function by CTLA-4 binding to CD80/CD86 on immature dendritic cells, leading to increases in tryptophan catabolism and downregulation of T effector activity^{14,15,26}. Non-T cells including placental fibroblasts also express CTLA-4 (ref. 30). Thus, the effects of alterations of the splicing of *CTLA4* might be wide ranging. Our results provide insights that focus attention on the sCTLA-4 isoform and its little-studied role in the maintenance of immune tolerance and autoimmunity. sCTLA-4 is known to be translated, secreted and present in human serum^{19–21}. It can bind the CD80/86 molecules, and recombinant sCTLA-4 inhibits T-cell proliferation *in vitro*²¹. The reduction in the level of sCTLA-4 mRNA produced by the disease-susceptible haplotype of *CTLA4* could lead to reduced blocking of CD80/CD86, causing increased activation through CD28, or to less stimulation of CD80/CD86 on immature dendritic cells, resulting in less tryptophan catabolism. The published correlation between *in vitro* T-cell activation and CTLA-4 genotype^{31–33} could be explained by variation in levels of sCTLA-4.

In the NOD mouse we propose that *Idd5.1* is a SNP in exon 2 of *Ctla4* that controls the splicing efficiency of a liCTLA-4 form. We can be certain that this RNA is translated into a functional protein from two other experiments. Recent studies of the human CD28 gene^{34,35} demonstrated that a CD28 isoform (CD28i) equivalent to liCTLA-4 exists. As found for liCTLA-4, CD28i contains the signal peptide, the transmembrane domain and the cytoplasmic tail, and, owing to an alternative splicing mechanism different from the one that we have observed in this study for CTLA-4, is missing the ligand-binding domain, encoded by CD28 exon 2. CD28i, which exists as both a homodimer and as a noncovalent heterodimer with CD28, functions as a co-stimulatory signal amplifier in human T cells³⁵. Second, there is direct evidence that liCTLA-4 protein causes a negative signal in CTLA-4-negative T cells transfected with liCTLA-4 complementary DNA (L. Vijaykrishnan, V. J. Kuchroo, L.S.W. and L.B.P., unpublished data). liCTLA-4 would presumably act by inhibiting signalling in T-cell activation and expansion. The existence of this functional form and our genetic data provide support for a ligand-independent mode of action of CTLA-4.

It is possible that the common functional variants of the CTLA-4 gene have been the subject of evolutionary selection in host resistance to infectious disease. The autoimmune-disease susceptibility alleles could provide enhanced cellular immune responses for particular viral or intracellular infections, whereas the autoimmune-resistance alleles may augment humoral immune responses against a different set of pathogens, such as parasitic worms. Immunoglobulin- ϵ is a major feature of the anti-worm response and also of atopic illness such as allergy and asthma. The *CTLA4* autoimmune-protective haplotype has been associated recently with susceptibility to atopy³⁶. It has been proposed that the significant increases in the incidence of T1D³⁷ and atopic

illness³⁸ in children over the last 40 years in affluent western populations might be due, in part, to advances in healthcare and in the decrease in frequency of worm and other infections. Taken together with our results we suggest that common allelic variation of *CTLA4* and other genes encoding regulators of the immune system may underpin the molecular basis for such gene–environment interactions. Our results point to a primary role in autoimmune disease for the alternative splice forms of CTLA-4—the soluble isoform in humans and the newly discovered ligand-independent form in mice. These observations highlight the usefulness of the positional genetics approach in discovering insights into disease mechanisms in a complex, multifactorial disorder, even for a gene as well studied as *CTLA4*. □

Methods

(See Supplementary Information for detailed Methods)

Subjects

White, UK-born Graves' disease cases ($n = 672$), Graves' disease family members ($n = 210$) and AIH cases ($n = 228$) were recruited from thyroid clinics in Birmingham (Queen Elizabeth and Heartlands Hospitals), Exeter (Royal Devon and Exeter Hospital) and Bournemouth (Royal Bournemouth Hospital), as described previously⁸. Ethnically matched control subjects ($n = 462$) with no history of autoimmune disease were recruited at various sites including the Blood Transfusion Service, Birmingham Heartlands Hospital and the Queen Elizabeth Hospital, Birmingham, and 382 control subjects (for which there are cell lines available at the European Collection Animal Cell Cultures) originated from blood donors from Oxford and Birmingham Blood Transfusion Services. T1D families were white European or of white European descent, with two parents and at least one affected child: 472 Diabetes UK Warren 1 multiplex from the United Kingdom³⁹; 322 United States multiplex from the Human Biological Data Interchange⁴⁰; 80 Yorkshire simplex from the United Kingdom; 32 southwest simplex from the United Kingdom; 216 Bristol simplex from the United Kingdom; 263 Belfast multiplex/simplex⁴¹ from the United Kingdom; 360 Norwegian simplex; 423 Romanian simplex; and 1,503 Finnish multiplex/simplex⁴². The Cambridge Local Research Ethics Committee approved experiments involving human subjects and informed consent was obtained from all subjects. The mouse research was approved by the Home Office and Cambridge University Biomedical Support Services.

SNP identification and genotyping

SNPs were identified in PCR products from 32 individuals (Supplementary Information A). SNPs were mainly genotyped using the Invader assay (Third Wave Technologies) in two different formats, uniplex⁴³ and biplex, which was validated for the present study (Supplementary Information A) and by others⁴⁴. The accuracy of genotyping data, its management and quality control are detailed in Supplementary Information A.

Statistical analyses

All statistical analysis, unless otherwise stated, was performed using STATA (<http://www.stata.com>). Particular use was made of the 'GenAssoc' routine, available from <http://www-gene.cimr.cam.ac.uk/clayton/software/stata/> (see also Supplementary Information B).

Gene expression

Peripheral blood mononuclear cells (PBMCs) were isolated from heparinized blood of healthy volunteers by density gradient centrifugation using Lympholyte H (Cedarlane Laboratories). CD4 T cells were purified by negative selection using magnetic beads (Dynal Biotech ASA). As determined by flow cytometry (FACS Scan and CELLQuest Software), cells were routinely >92–94% CD4⁺ and <1.5% CD8⁺. ASTQ²² was carried out on β CTLA-4 and sCTLA-4 cDNA products that had been amplified by PCR, with [α -³²P]dCTP incorporated in the final PCR step (Fig. 2b). cDNA amplification was performed using the primers: β CTLA-4, 5'-ACTTCCTGAAGACCTGAACACCG-3', 5'-GCACGGTCTCTGGATCAATTACA-3'; sCTLA-4, 5'-CTGAAGACCTGAACACCGCT-3', 5'-GGCTCTCTTTCTTAGCAATTACATA-3'. Cycling conditions were 95 °C for 25 s, 55 °C (sCTLA-4) or 60 °C (β CTLA-4) for 25 s and 72 °C for 25 s. Final cycle labelling with [α -³²P]dCTP was performed after 30 cycles of PCR. ASTQ PCR products were digested with *BbvI* and resolved by 8% polyacrylamide gel electrophoresis. Gels were dried and bands quantified after film exposure using the Quantity One software (BIO-RAD). The bands sometimes resolved into doublets, all of which were included in the quantitation. For the β CTLA-4/sCTLA-4 ratio, a real-time quantitative (Taqman) PCR assay was used employing the $\beta 2$ microglobulin gene as a control transcript and RNA purified from CD4 T cells (Fig. 2c). Relative levels were calculated from the ratio of the sCTLA-4 cycle threshold (CT) to β CTLA-4 CT values both of which were normalized to $\beta 2$ -microglobulin mRNA using the formula, $2^{(\Delta CT(sCTLA-4) - \Delta CT(\beta CTLA-4))}$, with ΔCT representing the difference in CT values between the target and control transcripts. Twelve of the volunteers in Fig. 2c had been previously tested in three additional experiments involving a smaller number of subjects in each ($n = 7, 4$ and 5). Aside from the two previously untested donors present in the fourth experiment shown in Fig. 2c, all samples have been replicated once with an inter-assay replication rate of 92%. A new cDNA synthesis of RNA was performed for each assay. Across all experiments the observed 50%

increase in the sCTLA-4/fCTLA-4 ratio in disease-protected CT60 A/A-genotype-positive individuals compared with susceptible G/G subjects was reproducible. In Fig. 3, primers (Supplementary Table A7) were designed across unique exon boundaries for fCTLA-4 (exons 2/3; 78 bp), sCTLA-4 (exons 2/4; 72 bp) and liCTLA-4 (exons 1/3; 113 bp). All expression differences were evaluated using a two-tail Student's *t*-test distribution.

Received 23 December 2002; accepted 3 April 2003; doi:10.1038/nature01621.

Published online 30 April 2003.

- Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genet.* **33**, 177–182 (2003).
- Dahlman, I. *et al.* Parameters for reliable results in genetic association studies in common disease. *Nature Genet.* **30**, 149–150 (2002).
- Lesage, S. & Goodnow, C. C. Organ-specific autoimmune disease: a deficiency of tolerogenic stimulation. *J. Exp. Med.* **194**, F31–F36 (2001).
- Finger, E. & Bluestone, J. When ligand becomes receptor-tolerance via B7 signaling on DCs. *Nature Immunol.* **3**, 1056–1057 (2002).
- Egen, J. G., Kuhns, M. S. & Allison, J. P. CTLA-4: new insights into its biological function and use in tumour immunotherapy. *Nature Immunol.* **3**, 611–618 (2002).
- Nistico, L. *et al.* The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum. Mol. Gen.* **5**, 1075–1080 (1996).
- Marron, M. P. *et al.* Genetic and physical mapping of a type 1 diabetes susceptibility gene (IDDM12) to a 100-kb phagemid artificial chromosome clone containing D2S72-CTLA4-D2S105 on chromosome 2q33. *Diabetes* **49**, 492–499 (2000).
- Allahabadia, A. *et al.* The MHC class II region, CTLA-4 gene and ophthalmopathy in patients with Graves' disease. *Lancet* **358**, 984–985 (2001).
- Kristiansen, O. P., Larsen, Z. M. & Pociot, F. CTLA-4 in autoimmune diseases—a general susceptibility gene to autoimmunity? *Genes Immun.* **1**, 170–184 (2000).
- Colucci, F., Bergman, M., Penha-Goncalves, C., Cilio, C. M. & Holmberg, D. Apoptosis resistance of nonobese diabetic peripheral lymphocytes linked to the *Idd5* diabetes susceptibility region. *Proc. Natl Acad. Sci. USA* **94**, 8670–8674 (1997).
- Hill, N. J. *et al.* The NOD *Idd5* locus controls insulinitis and diabetes and overlaps the orthologous CTLA4/IDDM12 and *NRAMP1* loci in humans. *Diabetes* **49**, 1744–1747 (2000).
- Lamhamed-Cherradi, S. E. *et al.* Further mapping of the *Idd5.1* locus for autoimmune diabetes in NOD mice. *Diabetes* **50**, 2874–2878 (2001).
- Sharpe, A. H. & Freeman, G. J. The B7-CD28 superfamily. *Nature Rev. Immunol.* **2**, 116–126 (2002).
- Grohmann, U. *et al.* CTLA-4-Ig regulates tryptophan catabolism *in vivo*. *Nature Immunol.* **3**, 1097–1101 (2002).
- Lin, H. *et al.* Cytotoxic T lymphocyte antigen 4 (CTLA4) blockade accelerates the acute rejection of cardiac allografts in CD28-deficient mice: CTLA4 can function independently of CD28. *J. Exp. Med.* **188**, 199–204 (1998).
- Chikuma, S., Imboden, J. & Bluestone, J. A. Negative regulation of T cell receptor-lipid raft interaction by cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* **197**, 129–135 (2003).
- Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- Cordell, H. J. & Clayton, D. G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am. J. Hum. Genet.* **70**, 124–141 (2002).
- Magistrelli, G. *et al.* A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. *Eur. J. Immunol.* **29**, 3596–3602 (1999).
- Oaks, M. K. & Hallett, K. M. Cutting edge: a soluble form of CTLA-4 in patients with autoimmune thyroid disease. *J. Immunol.* **164**, 5015–5018 (2000).
- Oaks, M. K. *et al.* A native soluble form of CTLA-4. *Cell Immunol.* **201**, 144–153 (2000).
- Kaijzel, E. L. *et al.* Allele-specific quantification of tumour necrosis factor alpha (TNF) transcription and the role of promoter polymorphisms in rheumatoid arthritis patients and healthy individuals. *Genes Immun.* **2**, 135–144 (2001).
- Lynch, K. W. & Weiss, A. A CD45 polymorphism associated with multiple sclerosis disrupts an exonic splicing silencer. *J. Biol. Chem.* **276**, 24341–24347 (2001).
- Zheng, Z.-M., He, P.-J. & Baker, C. Function of a bovine papillomavirus type 1 exonic splicing suppressor requires a suboptimal upstream 3' splice site. *J. Virol.* **73**, 29–36 (1999).
- Asano, M., Toda, M., Sakaguchi, N. & Sakaguchi, S. Autoimmune disease as a consequence of developmental abnormality of a T cell population. *J. Exp. Med.* **184**, 387–396 (1996).
- Manzotti, C. *et al.* Inhibition of human T cell proliferation by CTLA-4 utilizes CD80 and requires CD25⁺ regulatory cells. *Eur. J. Immunol.* **32**, 2888–2896 (2002).
- Maloy, K. J. *et al.* CD4 + CD25 + T(R) cells suppress innate immune pathology through cytokine-dependent mechanisms. *J. Exp. Med.* **197**, 111–119 (2003).
- Sakaguchi, S. Control of immune responses by naturally arising CD4(+) regulatory T cells that express Toll-like receptors. *J. Exp. Med.* **197**, 397–401 (2003).
- Pasare, C. & Medzhitov, R. Toll pathway-dependent blockade of CD4⁺CD25⁺ T cell-mediated suppression by dendritic cells. *Science* **299**, 1033–1036 (2003).
- Kaufman, K. A. *et al.* The CTLA-4 gene is expressed in placental fibroblasts. *Mol. Hum. Reprod.* **5**, 84–87 (1999).
- Huang, D., Gisco, R., Zhou, Y., Pirskanen, R. & Lefvert, A. Dinucleotide repeat expansion in the CTLA-4 gene leads to T cell hyper-reactivity via the CD28 pathway in myasthenia gravis. *J. Neuroimmunol.* **105**, 69–77 (2000).
- Kouki, T. *et al.* CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease. *J. Immunol.* **165**, 6606–6611 (2000).
- Maurer, M. *et al.* A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation. *Immunogenetics* **54**, 1–8 (2002).
- Magistrelli, G. *et al.* Identification of three alternatively spliced variants of human CD28 mRNA. *Biochem. Biophys. Res. Commun.* **259**, 34–37 (1999).
- Hanawa, H. *et al.* A novel costimulatory signalling in human T lymphocytes by a splice variant of CD28. *Blood* **99**, 2138–2145 (2002).
- Howard, T. *et al.* Fine mapping of an IgE-controlling gene on chromosome 2q: analysis of CTLA4 and CD28. *J. Allergy Clin. Immunol.* **110**, 743–751 (2002).
- Todd, J. A. A protective role of the environment in the development of type 1 diabetes? *Diabetic Med.* **8**, 906–910 (1991).
- Strachan, D. P. Hay fever, hygiene, and household size. *Br. Med. J.* **299**, 1259–1260 (1989).
- Bain, S. C., Todd, J. A. & Barnett, A. H. The British Diabetic Association — Warren Repository. *Autoimmunity* **7**, 83–85 (1990).
- Lernmark, A. *et al.* Family cell lines available for research. *Am. J. Hum. Genet.* **47**, 1028–1030 (1990).
- Patterson, C. C., Carson, D. J. & Hadden, D. R. Epidemiology of childhood IDDM in Northern Ireland 1989–1994: low incidence in areas with highest population density and most household crowding. Northern Ireland Diabetes Study Group. *Diabetologia* **39**, 1063–1069 (1996).
- Tuomilehto, J. *et al.* Epidemiology of childhood diabetes mellitus in Finland—background of a nationwide study of type 1 (insulin-dependent) diabetes mellitus. The Childhood Diabetes in Finland (DiMe) Study Group. *Diabetologia* **35**, 70–76 (1992).
- Mein, C. A. *et al.* Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res.* **10**, 330–343 (2000).
- Olivier, M. *et al.* High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res.* **30**, e53 (2002).
- Clayton, D. In *Handbook of Statistical Genetics* (eds Balding, D., Bishop, M. & Cannings, C.) 519–540 (Wiley, Chichester, 2001).
- Spielman, R., McGinnis, R. & Ewens, W. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Juvenile Diabetes Research Foundation (JDRF), the Wellcome Trust, NovoNordisk, Novo Nordisk Foundation, the Academy of Finland, The Sigrid Juselius Foundation, Diabetes UK and the Medical Research Council for financial support. H.U. was a Wellcome Trust Travelling Fellow and R.N. is a West Midlands Regional Health Authority Sheldon Medical Research Fellow. The availability of NOD congenic mice through the Taconic Farms Emerging Models Program has been made possible and is supported by grants from the Merck Genome Research Institute, NIAID and the JDRF. We gratefully acknowledge the participation of all patients, controls and family members, including provision of samples from T1D families from the Human Biological Data Interchange and Diabetes UK repositories, and sample collections by The Norwegian Study Group for Childhood, preparation of manuscript materials by J. Brown and T. Thorn, and review of the manuscript by C. Rudd, C. Nathan, M. Bobrow, P. Lyons, R. Glynn, C. Goodnow, J. Trowsdale, N. Proudfoot, T. Merriman, N. Hastie and D. Sansom.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.T. (john.todd@cimr.cam.ac.uk) or L.S.W. (linda.wicker@cimr.cam.ac.uk).

Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen

Dawei Li^{*†¶}, Rui Zhao^{*†}, Wayne Lilyestrom^{*}, Dahai Gai^{*}, Rongguang Zhang[‡], James A. DeCaprio[§], Ellen Fanning^{||}, Andrzej Jochimiak[‡], Gerda Szakonyi^{*} & Xiaojiang S. Chen^{*}

^{*} Department of Biochemistry and Molecular Genetics, University of Colorado Health Science Center, School of Medicine, Denver, Colorado 80262, USA

[‡] Advanced Photon Source, SBC, Argonne National Laboratory, Argonne, Illinois 60439, USA

[§] Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

^{||} Biological Sciences, VU Station B 1634, Vanderbilt University, Nashville, Tennessee 37235, USA

[†] These authors contributed equally to this work.

[¶] Present address: National Laboratory, China Agriculture University, Beijing 100049, China

The oncoprotein large tumour antigen (LTag) is encoded by the DNA tumour virus simian virus 40. LTag transforms cells and induces tumours in animals by altering the functions of tumour suppressors (including pRB and p53) and other key cellular proteins. LTag is also a molecular machine that distorts/melts the replication origin of the viral genome and unwinds duplex DNA. LTag therefore seems to be a functional homologue of the eukaryotic minichromosome maintenance (MCM) complex. Here we present the X-ray structure of a hexameric LTag with DNA helicase activity. The structure identifies the p53-binding surface and reveals the structural basis of hexamerization. The hexamer contains a long, positively charged channel with an unusually large central chamber that binds both single-stranded and double-stranded DNA. The hexamer organizes into two tiers that can potentially rotate relative to each other through connecting α -helices to expand/constrict the channel, producing an 'iris' effect that could be used for distorting or melting the origin and unwinding DNA at the replication fork.

LTag, an oncoprotein encoded by polyomaviruses (reviewed in ref. 1), has diverse biological functions and also performs essential functions in replicating the viral DNA (reviewed in refs 2–4). During viral replication, LTag assembles at the origin of replication as a double hexamer that distorts and melts the origin DNA^{5–7}. The LTag double hexamer is also the replicative helicase that unwinds duplex DNA at the replication forks^{8–12}. The simian virus 40 (SV40) LTag is a 708-residue protein with multiple domains. Its helicase domain has been mapped to residues 131–627 (refs 13, 14), which contain a core-origin-binding domain (LTag-obd, residues 131–250)^{15,16} and three conserved motifs of the helicase superfamily III (SF3; ref. 17), which are also members of the AAA+ superfamily¹⁸.

Here we report the 2.8-Å crystal structure of a LTag fragment that assembles into a hexamer with helicase activity. This is the first structure for a hexameric helicase that belongs to the helicase SF3 and AAA+ superfamilies. The structure defines LTag hexamer assembly interactions and provides insight into LTag's role in cell transformation. Furthermore, multiple hexameric structures of LTag reveal conformational changes leading to the expansion and constriction of the central channel, suggesting a potential mechanism for origin melting and the unwinding of DNA at the replication fork.

Overall structure of LTag hexamer

The SV40 LTag structure contains residues 251–627 (LTag251–627). This LTag251–627 fragment has helicase activity (Fig. 1a), which is unexpected given that the minimal helicase domain was previously assigned to residues 131–627 (refs 13, 14). Hereafter we shall refer to LTag251–627 as the helicase domain, even though the fragment is also capable of transforming cells^{19,20} and actually contains three distinct structural domains.

The LTag helicase domain assembles into a hexameric structure with a central channel (Fig. 1b, c). Viewed from the side, the hexamer ring appears as two layers (or tiers) of different diameters, the smaller tier at the top and the larger tier at the bottom (Figs 1b

and 4a). Viewed from the top, the hexameric ring has six points radiating from the centre of the larger bottom tier (Fig. 1c). The amino-terminal domain located in the top tier reaches over clockwise to its neighbour, stacking on top of the larger carboxy-terminal domains of the neighbouring molecule. A cleft exists between each pair of monomers in the larger bottom tier (Fig. 1c), which might provide space for the hexamer to undergo an ATP-dependent conformational change, driving DNA remodelling.

A gap separates the top from the bottom tier (Fig. 1b); α -helical structure connects the smaller and the larger tiers. The thickness of the two-tiered hexamer is ~ 80 Å. Previous electron-microscope studies²¹ have shown that the thickness of a LTag hexamer is ~ 120 Å, indicating that the missing N terminus (residues 1–250) would extend ~ 40 Å (ref. 22) on top of the smaller tier. This dimension of LTag (~ 120 Å per hexamer, or ~ 240 Å per double hexamer) is similar to that of an archaeal MCM double hexamer²³ and suggests that a LTag double hexamer at the replication origin should cover ~ 70 -base-pair (bp) double-stranded DNA (dsDNA), which is consistent with the observation that LTag protects 74-bp (or about seven turns) dsDNA¹⁶. Similarly, *Xenopus* MCM protects ~ 80 -bp DNA²⁴.

LTag monomer structure

The LTag monomer contains 16 α -helices and 5 β -strands (Fig. 2c). These structural elements fold into three domains: D1 (domain 1), D2 and D3 (Fig. 2a, b). D1 is the Zn domain at the N terminus and contains five α -helices ($\alpha 1$ – $\alpha 5$). The Zn atom coordinated by a Zn motif is important in holding $\alpha 3$ (α -helix 3) and $\alpha 4$ together (Fig. 2a), which in turn provide an anchor for $\alpha 1$ and $\alpha 2$. The beginning of $\alpha 5$ packs with $\alpha 1$ and $\alpha 2$ of D1, but its C terminus extends to $\alpha 6$ of D3.

The D2 domain contains three conserved helicase motifs related to SF3 helicases, namely the modified version of Walker A and B motifs and a motif C (ref. 17). D2 folds into a core β -sheet

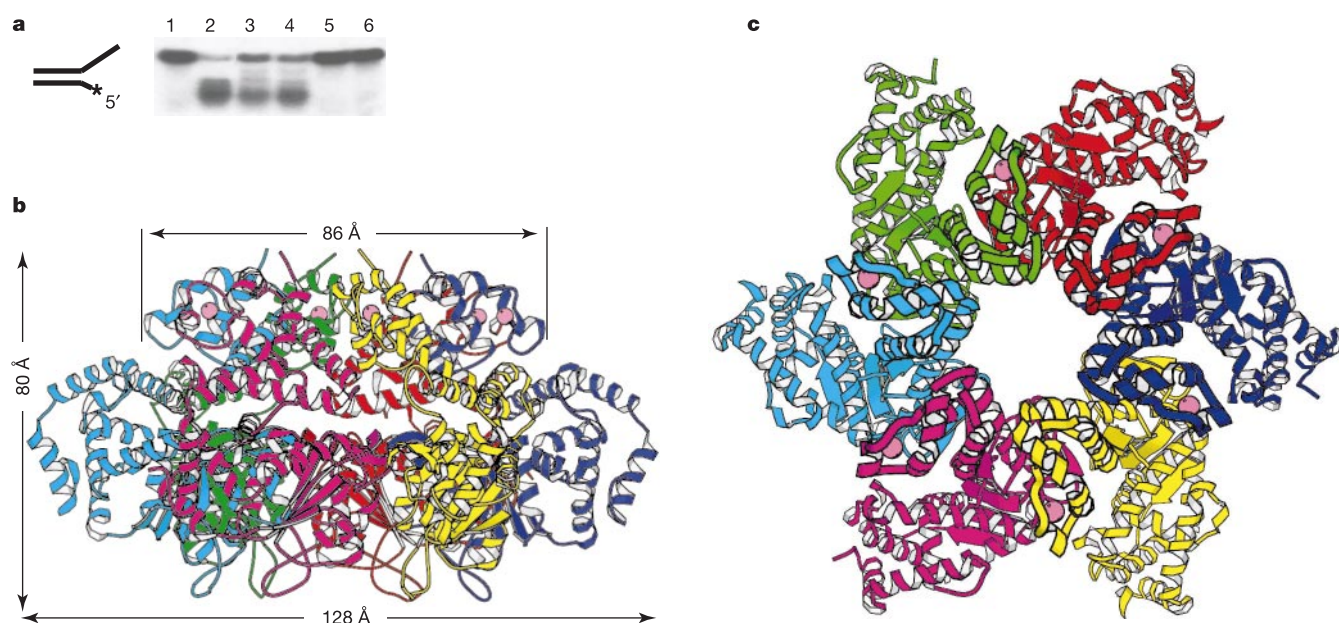


Figure 1 The helicase activity and hexamer structure of LTag251–627. **a**, The helicase assay was performed with a fork DNA substrate containing 20-bp dsDNA, a 16-nucleotide 3' overhang, and a three-nucleotide 5' overhang labelled with ^{32}P . All lanes contained 0.1 pmol DNA. Lanes 1 and 2, native and boiled DNA; lanes 3 and 4, 0.5 and 0.7 pmol LTag251–627 hexamer, respectively, showing helicase activity comparable to that of the full-length LTag (data not shown); lanes 5 and 6, 0.35 and 1.0 pmol LTag303–627, respectively, showing no helicase activity. **b**, **c**, Side (**b**) and top (**c**) views of the LTag251–627 hexameric structure. The pink balls represent Zn atoms. Each monomer is depicted with a distinct colour. Structures were generated with MOLSCRIPT⁴⁹.

consisting of five parallel β -strands sandwiched by α -helices (Fig. 2b). The five- β -stranded fold is similar to the AAA+ module of other AAA+ family members, such as NSF-D2 (membrane fusion²⁵), HslU (polypeptide unfolding and translocation^{26,27}) and RuvB (branch migration²⁸), all of which form hexamers. In addition to the five- β -stranded sheet, the LTag D2 AAA+ module has two antiparallel β -strands ($\beta 3'$ and $\beta 3''$; Fig. 2a, b) unique to LTag among the known structures of the AAA+ family. D2 also contains four helices, $\alpha 9$ – $\alpha 12$. Helices $\alpha 10$ – $\alpha 12$ are located on one side of the parallel β -sheet, and $\alpha 9$ is on the other (Fig. 2b). The loop between $\beta 1$ and $\alpha 9$ is the P-loop (Walker A motif), which functions to bind and orient the triphosphate of NTP for hydrolysis.

The third domain, D3, is all α -helical. Its seven α -helices originate from both the N-terminal region ($\alpha 6$ – $\alpha 8$) and the C terminus ($\alpha 13$ – $\alpha 16$), with D2 inserted between. One unique feature of the structure of D3 is that the three helices from the N-terminal region ($\alpha 6$ – $\alpha 8$) circle to form a nearly closed ring that is 'cross-linked' with another 'ring' composed of four α -helices at the C terminus ($\alpha 13$ – $\alpha 16$; Fig. 2a). The two 'crosslinked' α -helical rings of D3 pack tightly against D2– $\alpha 9$, the 'P-loop' helix important for ATP binding, forming a large globular bulge on one side of the D2 β -sheet (Fig. 2a). In contrast, D1 and D2 or D3 are well separated, connected by long helices ($\alpha 5$ and $\alpha 6$) that could allow motions between domains.

Hexamerization

The structure of the LTag hexamer was determined from the P321 crystal form obtained in the absence of ATP, in which one asymmetric unit contains two LTag molecules (Fig. 3a). The two molecules in one asymmetric unit are nearly identical, and the backbone can be superimposed with a root-mean-square deviation of 0.49 Å. Three dimers come together around the crystallographic 3-fold axis to form a hexamer (Fig. 1c), which should reflect the hexamer formed in solution in the absence of ATP (see next section). The monomer–monomer interactions within a hexamer come from two interfaces, the D1–D1 interface in the smaller tier (Fig. 3a, top) and the D2–D2

interface in the larger tier (Fig. 3a, bottom). D3 does not participate directly in the interactions at the interface. Rather, it constitutes the outer layer of the larger tier, giving the appearance of triangular points radiating from the centre (Fig. 1c).

The D1–D1 and D2–D2 interfaces, which bury a total surface area of 2,529 Å² between two monomers, involve two distinct types of interaction. The D1–D1 interface is uniformly hydrophobic, involving a leucine-zipper-like interaction between $\alpha 2$ of one molecule and $\alpha 5$ of its neighbour (Fig. 3b). Flanking the leucine-rich interface of $\alpha 2$ and $\alpha 5$, charged residues use their hydrophobic carbon side chains to pack against the hydrophobic residues on the neighbouring molecule. In contrast, the D2–D2 interactions are uniformly hydrophilic, involving hydrogen bonds and salt bridges through polar and charged residues (Fig. 3c). A total of 19 charged residues are present at this interface, plus several additional polar residues, including serine, asparagine and glutamine.

The role of the Zn motif and Zn domain

The Zn motif of LTag was proposed to form a canonical zinc-finger structure for DNA binding^{2,29}. However, the LTag251–627 structure reveals that the Zn domain (D1) has a globular fold (Fig. 2a) stabilized by the coordination of a Zn atom through the Zn motif (C302–C305–H313–H317) that deviates from the previously predicted motif (C302–C305–H317–H320; ref. 29), and no classical zinc-finger structure specialized for DNA binding is present. Furthermore, the helices coordinated by the Zn motif are located away from the central channel of the hexamer (Fig. 1c), which probably accommodates DNA. We therefore conclude that the Zn motif is not directly involved in binding DNA but is instead important for stabilizing the Zn-domain structure.

The structure of the LTag251–627 hexamer suggests that the Zn domain might be important for LTag hexamerization. To test this prediction, a deletion mutant containing residues 303–627 (LTag303–627) was constructed. LTag303–627 lacks $\alpha 1$ and $\alpha 2$ (important for D1–D1 interactions; Figs 2a and 3b) and C302 of the Zn motif, therefore lacking an intact Zn domain. The oligomeric

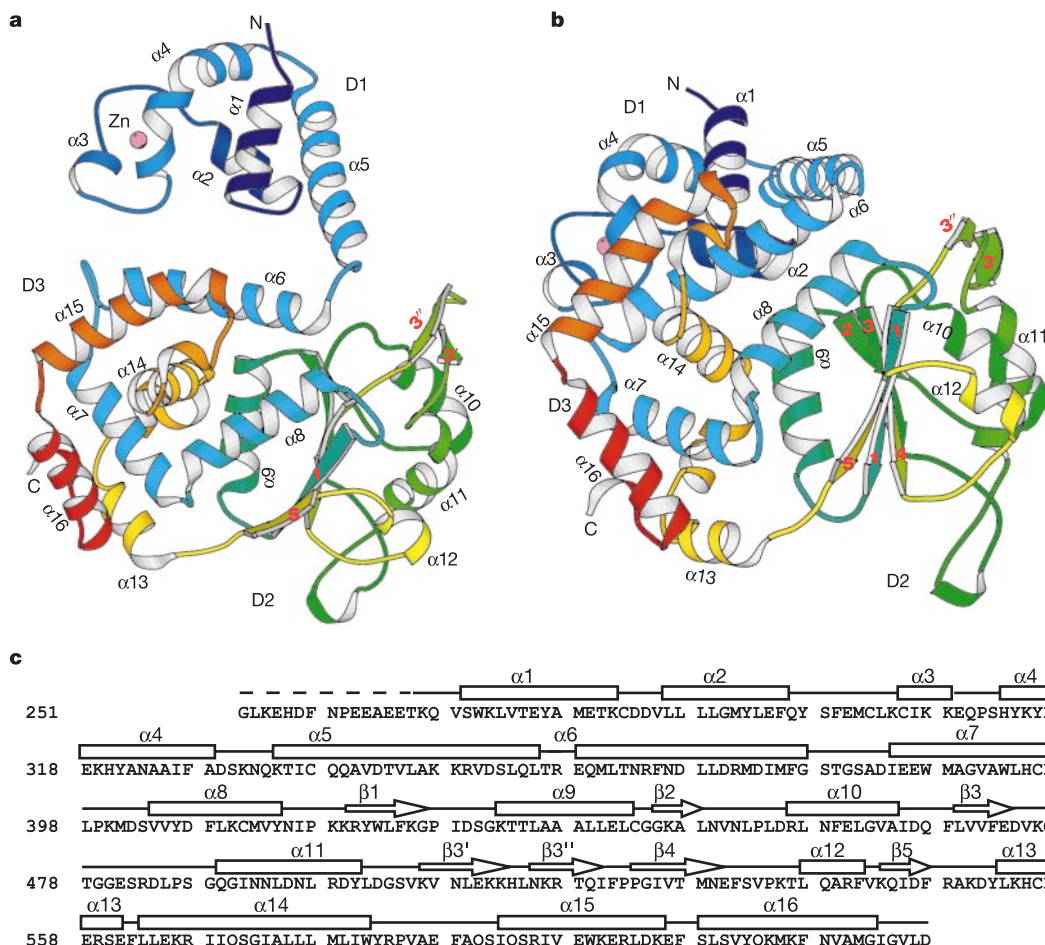


Figure 2 Detailed structure of LTag251–627 monomer. **a, b**, Two views of the LTag251–627 monomer that are related by a 60° rotation along the horizontal axis. The three domains D1, D2 and D3 are indicated. α -Helices and β -strands are numbered sequentially. **c**, The secondary structure elements of LTag. Open bars represent β -strands, arrows α -helices, solid lines coiled structure, and the dotted line represents a flexible region.

states of LTag303–627 and LTag251–627 were examined by gel-filtration chromatography. In the absence of ATP, LTag251–627 eluted in two peaks with an apparent molecular mass corresponding to hexamers and monomers in a 2:1 ratio (Fig. 3d). When the protein isolated from either the hexamer or the monomer peak was subjected to another round of chromatography, both the hexamer and monomer peaks reappeared in the same 2:1 ratio, indicating that LTag251–627 equilibrates between hexamers and monomer. However, if ATP was present, all protein was eluted in the hexamer peak (Fig. 3d). The full-length LTag behaved similarly to LTag251–627 (data not shown). This demonstrates that LTag251–627 contains all the elements needed for hexamerization and hexamerizes in the absence of ATP, even though such hexamers readily dissociate into monomers. ATP stimulates hexamerization and stabilizes the hexamers.

In contrast, LTag303–627, which contains D2 but lacks D1 (Zn domain), was eluted as one peak with an apparent molecular mass close to a monomer in gel-filtration chromatography, even in the presence of ATP (Fig. 3d). The monomeric state is consistent with its lack of ATPase activity (data not shown). Because ATP hydrolysis by LTag probably requires residues supplied *in trans* by a neighbouring molecule (see next section), only dimers or higher oligomers of LTag303–627 should possess ATPase activity. LTag303–627 therefore seems to be monomeric, indicating that the D2–D2 interactions in LTag303–627 are insufficient for dimer or hexamer formation with or without ATP, despite the large D2–D2 interface seen in the structure.

ATP binding by LTag hexamer

The hexamer structure in the p321 crystal form does not contain ATP. At the D2–D2 interface within a hexamer, three charged residues, Lys 418, Arg 498 and Asp 499, reside on one side of the interface with their side chains pointing towards the P-loop of the neighbouring monomer without making bonding contacts with any residues. The binding of ATP/ADP and Mg^{2+} at the P-loop of the neighbouring monomer should allow the three residues to make contact with either the ATP/ADP or the associated Mg^{2+} ion, which should strengthen the interactions between monomers to form the hexamer. These additional interactions augmented by ATP binding could enhance the cooperative process in LTag hexamerization and might explain why LTag251–627 equilibrates between hexamers and monomers in the absence of ATP but exists as stable hexamers in the presence of ATP (Fig. 3d).

Besides interacting with the ATP bound by the neighbouring molecule, the three residues (Lys 418, Arg 498 and Asp 499) might also have a role in ATP hydrolysis. The importance of *in trans* contribution of residues from adjacent monomers for hydrolysing NTPs has also been observed for other NTPases, such as T7 gp4 hexameric helicases and GTPase^{30–32}. In T7 gp4, an arginine residue (termed Arg ‘finger’) from a neighbouring molecule interacts with the γ -phosphate of ATP, permitting the efficient hydrolysis of ATP. Arg 498 of LTag occupies a position that might allow it to have an equivalent role to the Arg ‘finger’ for ATP hydrolysis.

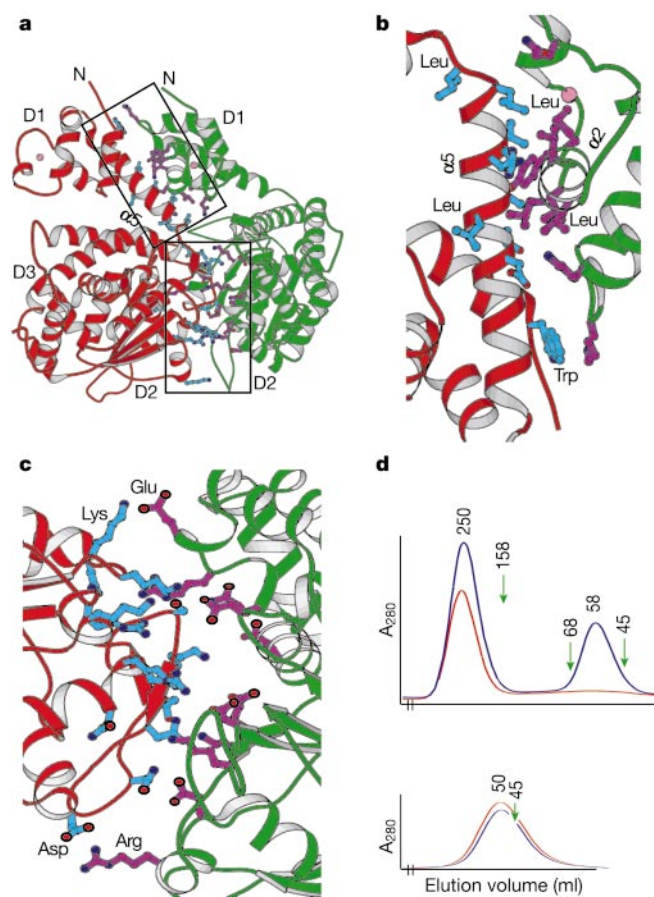


Figure 3 Hexamerization of LTag. **a**, An overview of the interface between two monomers (in red and green) in a hexamer, showing two separate interfaces (boxed regions): D1–D1 and D2–D2. **b**, The D1–D1 interface. $\alpha 5$ and $\alpha 2$ from two monomers pack against each other through their leucine-rich hydrophobic surfaces. **c**, The D2–D2 hydrophobic interface consisting of charged or polar side chains. **d**, The oligomeric states of LTag251–627 (upper) and LTag303–627 (lower) in solution. Both polypeptides were incubated in buffers without ATP (blue line) or with 1 mM ATP (red line) at 25 °C for 30 min before being loaded on a Superdex-200 column. The apparent molecular masses (kDa) of the peaks are indicated. The calculated molecular mass for LTag251–627 is 43.2 kDa per monomer (259.4 kDa per hexamer) and for LTag303–627 it is 37.2 kDa per monomer. The molecular masses and positions of standards (green arrows) are indicated.

Mutational inactivation in p53 binding and replication

The LTag structure provides a basis for understanding how particular mutations affect certain LTag functions. Mutations can be divided into three classes depending on whether they are defective in replication, in transformation or in both. The first class has mutations of residues in either D1 (including the Zn motif), affecting hexamerization, or in D2 and D3, disrupting ATPase activity and DNA replication^{29,33}. Most temperature-sensitive mutants belong to this class^{34,35}. The second class includes mutations of residues on the D3 surface that are important for the binding of p53 (including the four mutations P399L, D402E/N/H, C411R and P584L; Fig. 4a), disrupting cell transformation but not DNA replication^{19,36–38}. The four residues form a patch on the outer surface of a hexamer, suggesting that they either contact p53 directly or are positioned at the LTag–p53 interface. On the p53 side, mapping the residues important for binding LTag to the known p53 structure³⁹ shows an overlap within the DNA-binding surface of p53 (Fig. 4b, c), which is consistent with the observation that p53 binding by LTag prevents p53 from binding to its promoter, thus inhibiting p53 transactivation activity. The third class of mutations disrupts the D3 fold. D3 not only interacts with p53 and polymer-

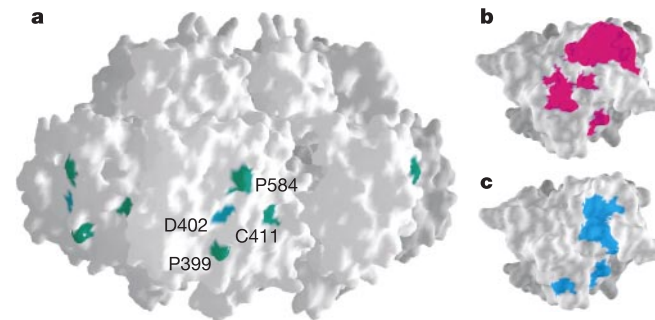


Figure 4 The p53-binding surface. **a**, Side view of LTag hexamer surface, showing a patch formed by the four residues (in cyan) important for binding p53. **b**, **c**, The surface of the p53 DNA-binding domain (p53–DBD)³⁹, showing the locations of residues important for binding LTag (magenta, in **b**), as well as for binding DNA (blue, in **c**). This shows that the two surfaces on p53–DBD overlap, indicating that LTag might inhibit p53 function by preventing its DNA-binding activity.

ase- α on its surface but also borders $\alpha 9$ and the P-loop (ATP-binding) of D2. Examples of this class of mutations are those with small deletions or insertions on $\alpha 14$ of D3 (Fig. 2a), which failed to replicate the viral DNA and lost their transformation activity⁴⁰.

Features and functions of LTag channel

Positively charged channels and the large ‘chamber’. Two essential functions of LTag in DNA replication are melting the origin and unwinding the duplex DNA at the fork. Hexameric LTag has a central channel whose inner surface is strongly positively charged (Fig. 5a), indicating a possible function in binding DNA, which is reminiscent of the DNA-binding channel of the archaeal MCM double hexamer²³. A new feature is the presence of a large ‘chamber’ in the middle of the channel (Fig. 5b), which is sufficiently wide (~ 67 Å) to accommodate strand separation during origin melting and fork DNA unwinding (see model in Fig. 6). Additionally, there are six positively charged holes formed between monomers on the side wall of the large ‘chamber’ (Fig. 5b). These holes (side channels) could potentially be used as outlets for the unwound single-stranded DNA (ssDNA) (see Fig. 6e).

DNA binding. The central-channel openings of the LTag251–627 hexamer differ for the two tiers (Fig. 5a). The smaller top tier has an opening of ~ 23 Å between side chains at the narrowest point, a dimension sufficient to allow the passage of dsDNA. The larger

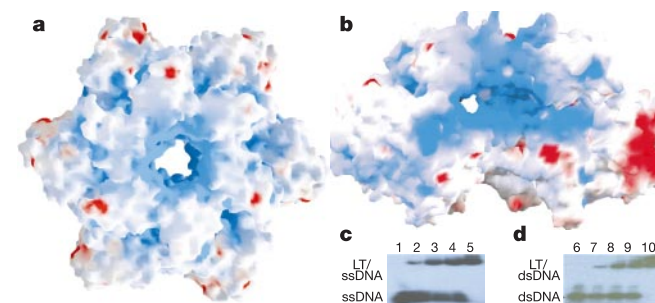


Figure 5 The channel features and DNA-binding activity of LTag hexamer. **a**, Surface representation of LTag hexamer with the smaller tier in front, showing the positively charged (blue) central channel. Red, negatively charged surface; white, uncharged surface (GRASP⁵⁰). **b**, The interior of the central channel, showing the large ‘chamber’ and the side channel. Two monomers have been taken away from the front to show the vertical path of the central channel, which widens inside the larger tier, forming a ‘chamber’. **c**, **d**, Autoradiographs of gels showing LTag (LT) hexamer binding to ssDNA (**c**) or dsDNA (**d**). Lanes 1–5 and lanes 6–10 contained 0, 0.1, 0.4, 1.6 and 6.4 pmol, respectively, LTag251–627 hexamer.

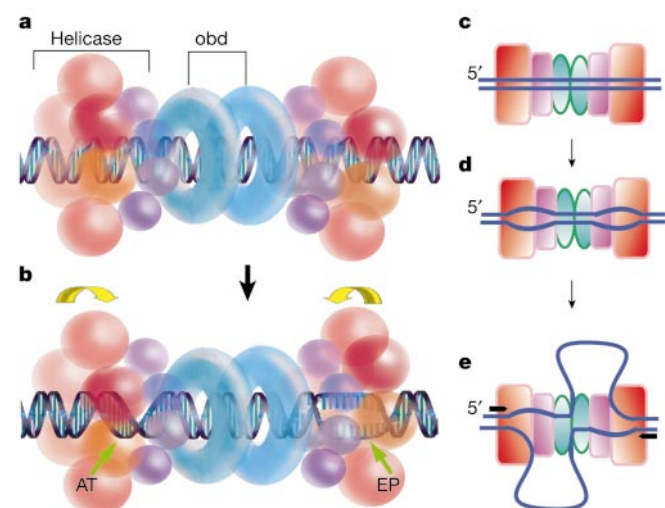


Figure 6 Models for origin distortion and melting, and DNA unwinding, by LTag. **a**, LTag double hexamer assembled at origin dsDNA. The helicase domain (LTag251–627) is labelled, with the larger ring representing the larger tier. **b**, The model for origin distortion and melting by the LTag helicase domain through the ‘iris’ mechanism. The distorted AT sequence and melted early palindrome (EP) region are indicated by arrows (green). **c–e**, A looping model for bidirectional DNA unwinding by LTag double hexamer. The colour coding of LTag domains is the same as in **a**. After LTag double hexamer assembly (**c**), the dsDNA is distorted and melted (**d**) by the ‘iris’ mechanism. One separated ssDNA strand is extruded through a side channel of the helicase domain to unwind the two forks bidirectionally (**e**).

bottom tier has an opening of ~ 15.0 Å (Fig. 5a), which is wide enough for ssDNA, but not dsDNA despite the large ‘chamber’ in the middle of this tier, which could accommodate two separated ssDNA strands simultaneously (Fig. 5b). DNA-binding assays show that LTag251–627 binds ssDNA (Fig. 5c), which is consistent with previous reports^{14,41}. The LTag251–627 hexamer also binds dsDNA (Fig. 5d), indicating that the LTag hexamer might undergo a structural rearrangement so that the narrowest openings of the central channel expand to accommodate dsDNA.

Observed changes of channel size. To examine whether the LTag hexamer can change channel openings, we determined the structures of two other LTag crystal forms obtained under different conditions. The channel openings of the two crystal forms as well as of the first crystal form (P321) differ significantly at the narrowest point. One structure has a narrower opening in the larger tier, 12.5 Å (compared with 15.0 Å in P321), but no change in the smaller tier, whereas the other structure has smaller channel openings for both tiers, with 10.1 Å (compared with 15.0 Å) for the larger tier and 20.1 Å (compared with 23 Å in P321) for the smaller tier at the narrowest point (Supplementary Figure). Significantly, the structures determined from individual crystals of the same crystal form grown under identical conditions, but soaked in different buffers, vary in channel openings by as much as 3.6 Å (data not shown). These observations suggest that the LTag hexamer is intrinsically dynamic, which is consistent with its role as a molecular machine. Examination of these structures reveals that the changes in channel dimensions are achieved partly through a slight untwist or twist between the smaller top tier and the larger bottom tier (Supplementary Figure). The untwist or twist motion between layers is also observed in the electron-microscope images of p97 hexamer⁴². Although binding to dsDNA can conceivably cause LTag to expand its central channel for dsDNA, these results show that the channel openings can expand or constrict even in the absence of dsDNA.

Channel expansion/constriction: the ‘iris’ motion. Three structural features of LTag hexamer might provide a basis for the expansion and constriction of the central channel in a similar

manner to the ‘iris’ (diaphragm) of a camera. First, each molecule has the D1 and the D2/D3 domains spanning the smaller and larger tiers in a twisted fashion, with D1 of one monomer extending into its neighbour (Fig. 1c), an arrangement potentially permitting an untwist or twist motion between domains embedded in the two separate tiers. Second, the connector between D1 and D2/D3, or between the smaller and large tiers, is a long α -helical structure ($\alpha 5$; Figs 1c and 2a), which might act like a spring allowing the two tiers to untwist and twist against each other. This untwist and twist motion between the tiers might cause the channel to expand and constrict like an ‘iris’. Third, the uniformly hydrophobic interactions at the D1–D1 interface and hydrophilic interactions at the D2–D2 interface (Fig. 3b, c) might allow the neighbouring monomers to slide against each other during the ‘iris’ motion in changing the hexameric channel openings.

Discussion

On the basis of the ‘iris’ hypothesis, we propose a model for the untwisting or melting of the origin dsDNA (Fig. 6a, b). It is known that a LTag double hexamer untwists or melts the AT-rich sequence on one side of the core origin and the early palindrome sequence on the other side, a process requiring ATP binding but not hydrolysis⁶. The dimensions of the LTag double hexamer binding to the core origin DNA¹⁵ imply that the two helicase domains at both ends must capture both the AT-rich and early palindrome sequences (Fig. 6a). We propose that the two tiers of the helicase domain untwist and twist relative to each other through the ‘iris’ mechanism, distorting the duplex DNA captured within the channel (Fig. 6b). Accompanying the untwisting of the two tiers is the expansion of the channel opening, which might facilitate the separation of the two strands of the distorted dsDNA (Fig. 6b), leading to the eventual melting of the origin for replication initiation. The wide ‘chamber’ (~ 67 Å in width) inside the channel (Fig. 5b) can provide sufficient space for strand separation of dsDNA.

Similarly, a repeated ‘iris’ motion, probably driven by ATP hydrolysis, could be employed for destabilizing dsDNA ahead of the fork for unwinding. Here we propose a model for the bidirectional DNA unwinding by a LTag double hexamer bound to the origin (Fig. 6c–e). First, LTag assembles at the origin by means of sequence-specific interactions mediated through LTag-obd (residues 131–250), forming a double hexamer that covers about seven turns of dsDNA (Fig. 6a, c). The helicase domain then uses the ‘iris’ mechanism to distort and melt the origin to initiate unwinding (Fig. 6d). Next, the LTag double hexamer unwinds dsDNA by means of the ‘iris’ mechanism at the two forks held together by the double hexamer (Fig. 6e), with dsDNA being pulled into the LTag double hexamer and the unwound ssDNA loop being extruded from the side wall of the double hexamer, a process probably powered by ATP hydrolysis.

The positively charged side channel (Fig. 5b) might provide an avenue for the exit of the unwound ssDNA loop (Fig. 6e). The exit of ssDNA through the side channel might have important implications. First, each LTag hexamer of a double hexamer would cover both ssDNA and dsDNA at a fork (Fig. 6e), which is consistent with previous DNase protection results^{10,43}. Second, the contact with dsDNA ahead of the fork, as well as the holding of a ssDNA strand in one of the six side channels behind the fork, would allow the relative rotation between the two tiers (‘iris’ mechanism) within each LTag hexamer to operate in untwisting the dsDNA. This configuration permits each hexamer in a functional double-hexamer unit^{5–7} to unwind two growing forks bidirectionally and eliminates the need for the two hexamers of a double-hexamer unit to track separately along (or rotate around) the helical path of dsDNA in opposite directions. Third, the exit of ssDNA from the side channel suggests that a complete opening of the hexamer ring is not required for the release of ssDNA, although a partial opening between two N-terminal obd domains would be needed to allow the unwound ssDNA to pass and reach to the side channel of the helicase domain.

The 'iris' model presented here describes a molecular event for origin distortion and melting. Furthermore, our model proposes that the helicase domain of a LTag double hexamer distorts and unwinds the duplex DNA bidirectionally using the 'iris' mechanism, and that the unwound ssDNA loops are released from the side channel of the helicase domains, a hypothesis consistent with the 'rabbit ear' observed by electron microscopy⁴³. The evidence that cellular MCM might also function as a double hexamer²³ and that the cellular DNA replication occurs at defined foci on the nuclear matrix where replication proteins are assembled⁴⁴ suggests that a mechanism similar to that proposed in the LTag models might exist for DNA replication in cellular chromosomes. □

Methods

Protein purification, characterization and crystallization

Crystallized LTag251–627 was produced in an *Escherichia coli* expression system as a glutathione *S*-transferase (GST) fusion protein. The protein was first purified on a glutathione affinity column. After cleavage of the GST fusion with thrombin, LTag was further purified by Superdex-200 gel-filtration chromatography. The protein was concentrated to 20 mg ml^{−1} in a buffer containing 25 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, 10 mM dithiothreitol. Crystals were grown at 4 °C by the hanging-drop vapour diffusion method. The P321 crystal grew from 20 mM dithiothreitol, 50 mM HEPES, pH 7.6, 105 mM NaCl, 5% ethanol, 1 mM EDTA. The cell constants for the P321 crystal are $a = b = 120.3$ Å, $c = 133.2$ Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$. Two other crystal forms were obtained at different pH values: pH 7.25 to give a C2 crystal form ($a = 194.3$ Å, $b = 85.9$ Å, $c = 127.8$ Å, $\alpha = \gamma = 90^\circ$, $\beta = 128.8^\circ$) and pH 7.0 to give an R3 crystal form ($a = b = 168.3$ Å, $c = 129.0$ Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$).

Data collection, structure determination and refinement

All the data sets were collected at the SBC 191D beamline at the Advanced Photon Source (Chicago, Illinois). Crystals were flash-frozen in liquid nitrogen in crystallization buffer supplemented with 25% glycerol. Data were processed with HKL2000 (Table 1; ref. 45). The structure of LTag was determined by the multiple anomalous dispersion (MAD) method with the program SOLVE⁴⁶ using a Zn-MAD data set. A solvent flattening step with SOLVE based on the Zn-MAD phases yielded an electron density map containing regions of well-featured α -helices, which allowed the initial model building with the program O. Cyclic refinement and model building led to a final model that, after an individual *B*-factor refinement, has an R_{free} of 27.9% for the 30.0–2.8-Å resolution data (Table 1). For the determination of the structures of the other two crystal forms, initial attempts of molecular replacement (AMoRe⁴⁷, CCP4 program suite⁴⁸) using the complete LTag251–627 model did not give obvious solutions, probably owing to the different domain orientations of LTag in these crystal forms. However, the correct solutions were obtained by the molecular replacement method when the LTag model minus the Zn domain (D1) was used as the search model.

DNA-binding assay

A 33-nucleotide oligomer labelled with ³²P was used as a ssDNA substrate. To generate a dsDNA substrate, this labelled ssDNA was annealed to its complementary oligonucleotide. LTag was preincubated with 1 mM ATP so that LTag was in its hexameric form (see Fig. 3d). To perform the binding assay, a fixed amount (0.1 pmol) of ssDNA or dsDNA substrate was added to a solution containing 1 mM ATP, 50 mM Tris-HCl, pH 8.0, 200 mM NaCl and LTag251–627 protein in varied quantities (0.0, 0.1, 0.4, 1.6 and 6.4 pmol hexamer). After a 10-min incubation at 25 °C, the reaction mixture was loaded on a 6% native polyacrylamide gel for electrophoresis analysis. The gel was dried for autoradiography.

Helicase assay

The substrate for helicase assay was a short-fork DNA annealed from two oligonucleotides that had been purified with MonoQ column chromatography. The substrate DNA was incubated with LTag251–627 at 25 °C for 30 min in a buffer containing 50 mM HEPES, pH 7.5, 1 mM ATP, 3 mM MgCl₂, 1 mM dithiothreitol and 50 mM NaCl. The reaction was terminated by adding a stop solution containing 100 mM EDTA, 0.5% SDS and 50% glycerol. Samples were analysed on a 12% native polyacrylamide gel in 1 × Tris/borate/EDTA running buffer. The unwinding of the substrate DNA was detected by autoradiography.

Received 9 December 2002; accepted 25 March 2003; doi:10.1038/nature01691.

- Pipas, J. M. Common and unique features of T antigens encoded by the polyomavirus group. *J. Virol.* **66**, 3979–3985 (1992).
- Simmons, D. T. SV40 large T antigen functions in DNA replication and transformation. *Adv. Virus Res.* **55**, 75–134 (2000).
- Stillman, B. Smart machines at the DNA replication fork. *Cell* **78**, 725–728 (1994).
- Bullock, P. A. The initiation of simian virus 40 DNA replication *in vitro*. *Crit. Rev. Biochem. Mol. Biol.* **32**, 503–568 (1997).
- Mastrangelo, I. A. *et al.* ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. *Nature* **338**, 658–662 (1989).
- Borowiec, J. A. & Hurwitz, J. Localized melting and structural changes in the SV40 origin of replication induced by T-antigen. *EMBO J.* **7**, 3149–3158 (1988).
- Joo, W. S., Kim, H. Y., Purviance, J. D., Sreekumar, K. R. & Bullock, P. A. Assembly of T-antigen double

- hexamers on the simian virus 40 core origin requires only a subset of the available binding sites. *Mol. Cell. Biol.* **18**, 2677–2687 (1998).
- Dean, F. B. *et al.* Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication. *Proc. Natl Acad. Sci. USA* **84**, 16–20 (1987).
- Wold, M. S., Li, J. J. & Kelly, T. J. Initiation of simian virus 40 DNA replication *in vitro*: Large-tumor-antigen- and origin-dependent unwinding of the template. *Proc. Natl Acad. Sci. USA* **84**, 3643–3647 (1987).
- Smelkova, N. V. & Borowiec, J. A. Synthetic DNA replication bubbles bound and unwound with twofold symmetry by a simian virus 40 T-antigen double hexamer. *J. Virol.* **72**, 8676–8681 (1998).
- Tsurimoto, T., Melendy, T. & Stillman, B. Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin. *Nature* **346**, 534–539 (1990).
- Tsurimoto, T. & Stillman, B. Replication factors required for SV40 DNA replication *in vitro*. II. Switching of DNA polymerase α and δ during initiation of leading and lagging strand synthesis. *J. Biol. Chem.* **266**, 1961–1968 (1991).
- Wun-Kim, K. & Simmons, D. T. Mapping of helicase and helicase substrate-binding domains on simian virus 40 large T antigen. *J. Virol.* **64**, 2014–2020 (1990).
- Wu, C., Roy, R. & Simmons, D. T. Role of single-stranded DNA binding activity of T antigen in simian virus 40 DNA replication. *J. Virol.* **75**, 2839–2847 (2001).
- Borowiec, J. A., Dean, F. B., Bullock, P. A. & Hurwitz, J. Binding and unwinding—how T antigen engages the SV40 origin of DNA replication. *Cell* **60**, 181–184 (1990).
- Fanning, E. & Knippers, R. Structure and function of simian virus 40 large tumor antigen. *Annu. Rev. Biochem.* **61**, 55–85 (1992).
- Koonin, E. V. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* **21**, 2541–2547 (1993).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Beachy, T. M., Cole, S. L., Cavender, J. F. & Tevethia, M. J. Regions and activities of simian virus 40 T antigen that cooperate with an activated *ras* oncogene in transforming primary rat embryo fibroblasts. *J. Virol.* **76**, 3145–3157 (2002).
- Cavender, J. F., Conn, A., Epler, M., Lacko, H. & Tevethia, M. J. Simian virus 40 large T antigen contains two independent activities that cooperate with a *ras* oncogene to transform rat embryo fibroblasts. *J. Virol.* **69**, 923–934 (1995).
- Valle, M., Gruss, C., Halmer, L., Carazo, J. M. & Donate, L. E. Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol. Cell. Biol.* **20**, 34–41 (2000).
- Luo, X., Sanford, D. G., Bullock, P. A. & Bachovchin, W. W. Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nature Struct. Biol.* **3**, 1034–1039 (1996).
- Fletcher, R., Bishop, B., Sclafani, R., Ogata, G. & Chen, X. The structure and function of MCM dodecamer from Archaeal *M. thermoautotrophicum*. *Nature Struct. Biol.* **10**, 160–167 (2003).
- Edwards, M. C. *et al.* MCM2–7 complexes bind chromatin in a distributed pattern surrounding ORC in *Xenopus* egg extracts. *J. Biol. Chem.* **277**, 33049–33059 (2002).
- Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weis, W. I. Crystal structure of the hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–536 (1998).
- Sousa, M. C. *et al.* Crystal and solution structures of an HslUV protease–chaperone complex. *Cell* **103**, 633–643 (2000).
- Bochtler, M. *et al.* The structures of HslU and the ATP-dependent protease HslU–HslV. *Nature* **403**, 800–805 (2000).
- Putnam, C. D. *et al.* Structure and mechanism of the RuvB Holliday junction branch migration motor. *J. Mol. Biol.* **311**, 297–310 (2001).
- Loeber, G., Parsons, R. & Tegtmeyer, P. The zinc finger region of simian virus 40 large T antigen. *J. Virol.* **63**, 94–100 (1989).
- Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. Crystal structure of the helicase domain from the replicative helicase–primase of bacteriophage T7. *Cell* **99**, 167–177 (1999).
- Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
- Scheffzek, K. *et al.* The Ras–RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1997).
- Farber, J. M., Peden, K. W. & Nathans, D. *Trans*-dominant defective mutants of simian virus 40 T antigen. *J. Virol.* **61**, 436–445 (1987).
- Loeber, G., Tevethia, M. J., Schwedes, J. F. & Tegtmeyer, P. Temperature-sensitive mutants identify crucial structural regions of simian virus 40 large T antigen. *J. Virol.* **63**, 4426–4430 (1989).
- Ray, S., Anderson, M. E., Loeber, G., McVey, D. & Tegtmeyer, P. Functional characterization of temperature-sensitive mutants of simian virus 40 large T antigen. *J. Virol.* **66**, 6509–6516 (1992).
- Lin, J. Y. & Simmons, D. T. The ability of large T antigen to complex with p53 is necessary for the increased life span and partial transformation of human cells by simian virus 40. *J. Virol.* **65**, 6447–6453 (1991).
- Lin, J. Y. & Simmons, D. T. Stable T–p53 complexes are not required for replication of simian virus 40 in culture or for enhanced phosphorylation of T antigen and p53. *J. Virol.* **65**, 2066–2072 (1991).
- Kierstead, T. D. & Tevethia, M. J. Association of p53 binding and immortalization of primary C57BL/6 mouse embryo fibroblasts by using simian virus 40 T-antigen mutants bearing internal overlapping deletion mutations. *J. Virol.* **67**, 1817–1829 (1993).
- Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–355 (1994).
- Peden, K. W., Srinivasan, A., Farber, J. M. & Pipas, J. M. Mutants with changes within or near a hydrophobic region of simian virus 40 large tumor antigen are defective for binding cellular protein p53. *Virology* **168**, 13–21 (1989).
- Wu, C., Edgill, D. & Simmons, D. T. The origin DNA-binding and single-stranded DNA-binding domains of simian virus 40 large T antigen are distinct. *J. Virol.* **72**, 10256–10259 (1998).
- Rouiller, I. *et al.* Conformational changes of the multifunctional p97 AAA ATPase during its ATPase cycle. *Nature Struct. Biol.* **9**, 950–957 (2002).
- Wessel, R., Schweizer, J. & Stahl, H. Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J. Virol.* **66**, 804–815 (1992).

44. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
45. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–325 (1997).
46. Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
47. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
48. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
49. Kraulis, P. E. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
50. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L.G. Chen for her assistance in the artwork, R. Garcea and L. Chen for comments on the manuscript, other members of the X. Chen laboratory for help and input, staff at 19id and 14bmc in Argonne National Laboratory and at X25 in Brookhaven National Laboratory for assistance in data collection, and the UCHSC X-ray centre in the Biomolecular Structure Program for support. This work is supported in part by start-up and cancer-centre funds from UCHSC to X.C. and NIH-R01 to X.C., J.A.D. and E.F., and a DOE grant to R.Z. and A.J.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.S.C. (xiaojiang.chen@uchsc.edu). The atomic coordinates have been deposited with the Protein Data Bank under accession number 1N25.

A class of compact dwarf galaxies from disruptive processes in galaxy clusters

M. J. Drinkwater*, M. D. Gregg†‡, M. Hilker§, K. Bekki||, W. J. Couch||
H. C. Ferguson¶, J. B. Jones# & S. Phillipps☆

* Department of Physics, University of Queensland, Queensland 4072, Australia

† Department of Physics, University of California, Davis, California 95616, USA

‡ Institute for Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, L-413, Livermore, California 94550, USA

§ Sternwarte der Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

|| School of Physics, University of New South Wales, Sydney 2052, Australia

¶ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

☆ Astrophysics Group, Department of Physics, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

Dwarf galaxies have attracted increased attention in recent years, because of their susceptibility to galaxy transformation processes within rich galaxy clusters^{1–3}. Direct evidence for these processes, however, has been difficult to obtain, with a small number of diffuse light trails⁴ and intra-cluster stars^{5,6} being the only signs of galaxy disruption. Furthermore, our current knowledge of dwarf galaxy populations may be very incomplete, because traditional galaxy surveys are insensitive to extremely diffuse or compact galaxies⁷. Aware of these concerns, we recently undertook an all-object survey of the Fornax galaxy cluster⁸. This revealed a new population of compact members^{9,10}, overlooked in previous conventional surveys. Here we demonstrate that these ‘ultra-compact’ dwarf galaxies are structurally and dynamically distinct from both globular star clusters and known types of dwarf galaxy, and thus represent a new class of dwarf galaxy. Our data are consistent with the interpretation that these are the remnant nuclei of disrupted dwarf galaxies, making them an easily observed tracer of galaxy disruption.

We used the Two Degree Field (2dF) multi-object spectrograph on the 3.9 m Anglo-Australian Telescope to make a large all-object spectroscopic survey of the Fornax galaxy cluster⁸. The 400-fibre multiplex advantage of the 2dF spectrograph allowed us to include unresolved objects (normally ignored as ‘stars’). Among these ‘stars’ we discovered^{9,10} a population of objects in the cluster with luminosities of $M_V \approx -11$ mag, intermediate between typical globular star clusters ($M_V \approx -8$ mag) and normal dwarf galaxies ($M_V \approx -11$ to -16 mag). These ‘ultra-compact’ dwarf (UCD) galaxies are unlike other dwarf galaxies of similar luminosity by virtue of their high central concentration and thus star-like morphology in typical one-arcsecond-resolution ground-based imaging. At the distance of the Fornax cluster¹¹ (20 Mpc; Hubble constant $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$) this implies sizes of, at most, 100 pc.

We proposed⁹ that the UCDs were either unusually large and isolated star clusters or a new type of compact galaxy, perhaps formed by the disruption of nucleated dwarf galaxies. The UCDs are more luminous than Galactic globular star clusters, but they could be the extreme bright tail of the populous globular cluster system around the central galaxy of the Fornax cluster NGC1399 (ref. 12). The UCD luminosities, however, also suggestively overlap the luminosity distribution of the nuclei of dwarf elliptical galaxies (which peaks at $M_V \approx -10$ mag; ref. 13). A third possibility is that UCDs are simply ordinary nucleated dwarf ellipticals, but at one extreme of a continuous sequence of envelope luminosities, having envelopes too faint to be easily detected. The lack of extended halos

around the UCDs in photographic images¹⁰ argued against this third hypothesis.

Here we present new observations of the UCDs, aimed at determining their nature. Our deep imaging of the Fornax cluster confirms that UCDs are morphologically distinct from ordinary dwarf ellipticals, and our high-resolution imaging and spectroscopy show that they are not globular star clusters; we therefore conclude that they represent a new class of dwarf galaxy. Our interpretation is that the UCDs are the stripped nuclei of dwarf elliptical galaxies and therefore represent a method of tracing galaxy disruption processes in clusters.

That normal dwarf galaxies and UCDs are quite distinct, morphologically, is shown in Fig. 1, where we present an envelope surface brightness versus core luminosity plot. Nucleated dwarf elliptical (dE,N) galaxies show a correlation between the effective surface brightness of their envelopes and their core luminosities. In contrast, UCDs lie in a completely different region of this diagram: because they have no detectable outer envelope, only upper limits can be put on their surface brightness, all of which lie 3.5 magnitudes (a factor of 25) fainter than the dE,N galaxies with the same nuclear luminosity. Any intermediate objects would easily have been detected in our survey, so UCDs are clearly not part of the normal distribution of dwarf elliptical galaxies. The UCDs are also more concentrated to the centre of the cluster than the dwarf galaxies⁹; the radial distribution of the seven known UCDs differs from that of the 37 dE,N galaxies in our survey at the 99.7% confidence level. We therefore conclude that UCDs are not extreme examples of normal dwarf galaxies with very faint envelopes.

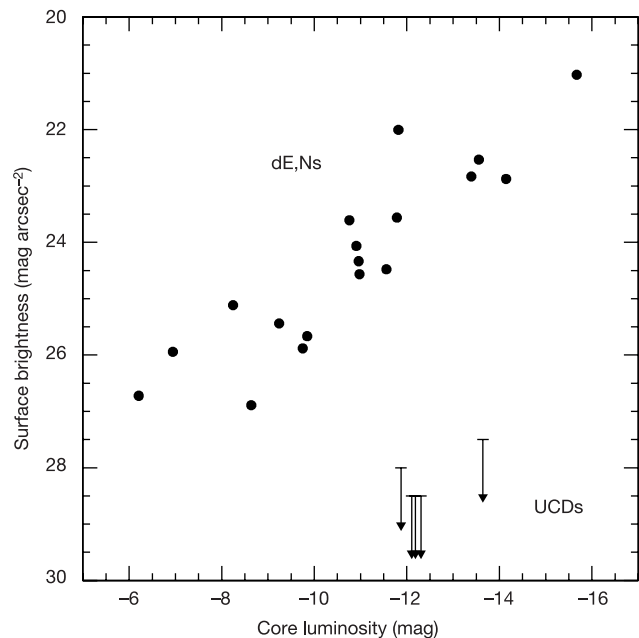


Figure 1 Comparison of the morphology of UCD galaxies with normal dE,N galaxies in the Fornax cluster. The envelope effective surface brightness in the V-band of the objects is plotted as a function of the V-band luminosity of their cores. The luminosity is in units of V-band absolute magnitude $M_V = \text{const.} - 2.5 \log_{10}(L)$, with luminosity L increasing from left to right. The surface brightness of the UCDs are upper limits, because no envelopes were detected on scales greater than the image resolution (2 arcseconds full-width half-maximum). The data are from deep images of 2.4 square degrees in the centre of the Fornax cluster taken with the 100-inch du Pont telescope of the Las Campanas Observatory²⁴. This region contains the original five UCDs as well as 18 dE,N galaxies²⁵.

We obtained high-resolution imaging with the Hubble Space Telescope (HST) of the first five UCDs discovered. We also observed a normal dE,N galaxy for comparison. Further comparisons are provided by similar published observations of six dwarf elliptical galaxies in the Virgo cluster¹⁴. Our images are shown in Fig. 2. The UCDs are very compact, typically only a few times larger than the 0.05-arcsecond image resolution. The UCD profiles were well fitted by both King models (normal for globular star clusters) and de Vaucouleurs $R^{1/4}$ law profiles (normal for giant elliptical galaxies) but are not consistent with the exponential profiles typical of normal dwarf galaxies (excluded at the 99.9% confidence level). The effective radii (defined to contain half the light) of the UCDs range from 10 to 22 pc. In the case of UCD3, the most luminous UCD, the model profile required an additional, but still very small (60-pc scale length), exponential halo component for a good fit. As

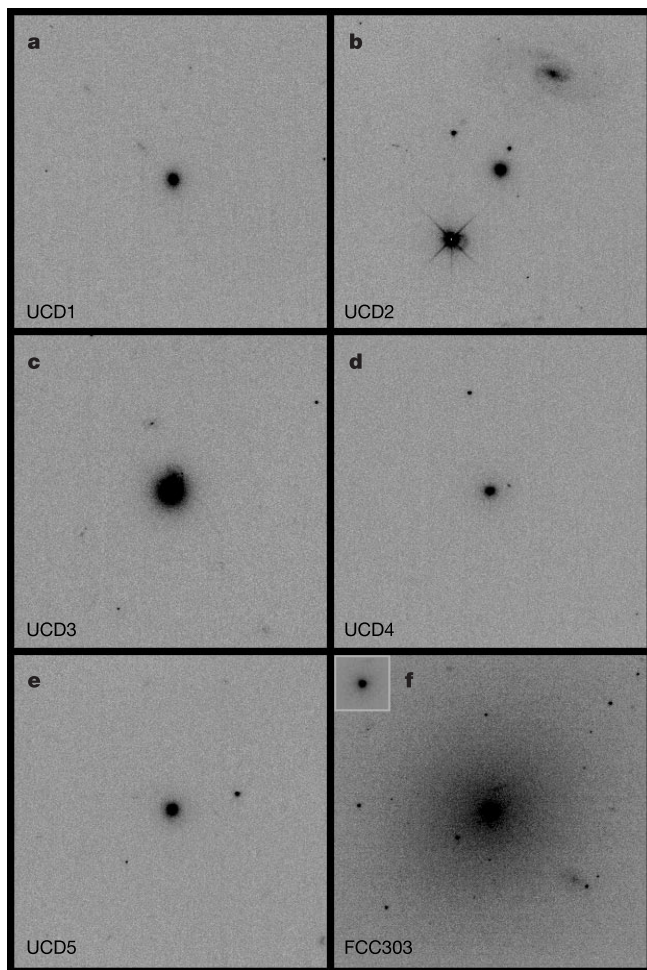


Figure 2 Comparison of Hubble Space Telescope images of the UCD galaxies with a normal nucleated dwarf galaxy (FCC303; ref. 25). Panel **f** is at the same contrast as **a–e**; the inset in **f** is at a lower contrast to show the nucleus. The extended halo of FCC303 more than fills the 30-arcsec field (2.9 kpc or 9.0×10^{19} m at the Fornax cluster). These images, shown in negative format, were obtained in the unfiltered mode of the Space Telescope Imaging Spectrograph (STIS). The one-orbit (40-min) exposures were made in the ‘clearpass’ optical CCD imaging mode of STIS for maximum sensitivity. We measured the UCD sizes with the ‘lshape’ program²⁶, which iteratively convolves a chosen model profile with the STIS point-spread function to match the galaxy profile by minimizing a χ^2 difference statistic. We used a fitting radius of 0.8 arcsec (16 pixels) corresponding to the same physical radius of 75 pc as used in the Virgo study¹⁴ and obtained effective radii for the UCDs of 10–22 pc.

expected, the comparison nucleated dwarf, FCC303, also required a two-component model of a core plus a larger (300-pc scale length) exponential halo. The UCD cores are all significantly larger than Galactic globular star clusters (typical effective radii of 3–5 pc) and are also larger than the 8-pc core of FCC303. Conversely, the UCDs are all smaller than known dwarf galaxies. Even the halo we do detect around UCD3 is tiny for a dwarf elliptical; the most compact normal dwarf galaxies in the Virgo cluster have scale lengths of 160 pc (ref. 15).

To study the internal dynamics and estimate the masses of the UCDs and FCC303, we measured their internal velocity dispersions with the Very Large Telescope (VLT) and the W.M. Keck Telescope. The velocity dispersions range from 24 to 37 km s^{-1} , considerably higher than those of Galactic globular star clusters, but overlapping the values published for the most luminous globular star clusters in the spiral galaxy M31 (ref. 16). Assuming the UCDs are dynamically relaxed systems, we can use their sizes and velocity dispersions to estimate their masses. Using a King model mass estimator¹⁷ for our profile fits we obtain masses of $1\text{--}5 \times 10^7 M_{\odot}$ (solar masses) for the UCDs and $1.4 \times 10^7 M_{\odot}$ for the dE,N galaxy nucleus, compared to around $4 \times 10^6 M_{\odot}$ for the most massive globular star clusters¹⁶. We obtained similar masses using Plummer models¹⁸, although these profiles did not fit out data as well. Using the HST data to measure the total V-band luminosities of the UCDs, we derived their mass-to-light ratios. The mass-to-light (M/L) ratios range from 2 to 4 in solar units, compared to around unity for typical globular star clusters. The largest globular star clusters in M31 have $M/L = 1\text{--}2$. In summary, the UCDs have luminosities, sizes and velocity dispersions similar to the cores of nucleated dwarf galaxies (in both the Fornax and Virgo¹⁴ clusters), but are larger and have higher mass-to-light ratios than even the largest globular star clusters.

The distinction between UCDs and other types of stellar system is made even clearer if we consider the dynamical correlations between the physical observables of luminosity and velocity dispersion shown in Fig. 3. The figure compares UCDs with globular clusters, dE,N galaxies and giant elliptical galaxies. We also include the large Milky Way globular cluster, ω Centauri, which is thought to have originated outside the Milky Way¹⁹. The UCDs lie well off the globular cluster $L \propto \sigma^{1.7}$ relation¹⁶, filling a previously unoccupied part of the diagram, but on an extrapolation of the elliptical galaxy $L \propto \sigma^4$ Faber–Jackson law²⁰.

The properties of the cores of the comparison dE,N galaxies are also shown: these are distributed between the globular clusters and the UCDs. The locations of the UCDs and the dwarf galaxy nuclei strongly support the ‘galaxy threshing’ hypothesis: UCDs are nucleated dwarf galaxies whose outer envelopes have been tidally stripped by interactions with the central cluster galaxy³. Removing the halo of a normal dE,N galaxy reduces the total luminosity by a factor of about 100 (5 mag), but barely changes the central velocity dispersion and nuclear luminosity (the dynamical relaxation time of the remaining core can be as long as 10^{11} years). The predicted effect of threshing is shown as a dotted line in Fig. 3 connecting the positions of FCC303 and that of its nucleus alone.

The case for tidal stripping is further supported by our detection of a small halo around the largest UCD. This object may be partially stripped, caught at an intermediate stage between UCDs and dE,N galaxies. As threshing will have been at least as important at earlier times in the life of the cluster, this implies a much larger population of UCDs than the seven currently known, although if they have the same luminosity function as dwarf galaxy nuclei¹³, most will be fainter than the limit of our survey.

Our new observations confirm that UCDs are qualitatively

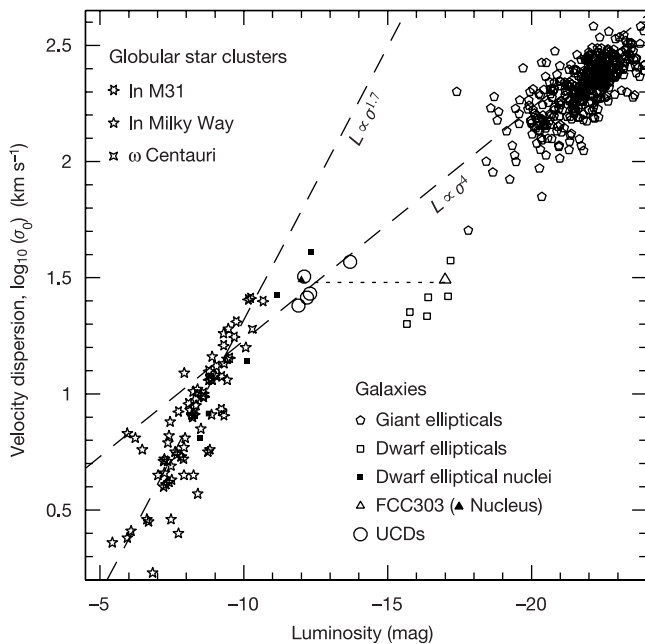


Figure 3 Comparison of the internal dynamics of the UCDs with normal galaxies and globular star clusters. The internal velocity dispersions (σ_0) of the different objects are plotted as a function of luminosity as in Fig. 1. The dashed lines show the Faber–Jackson relation²⁰ for elliptical galaxies and the steeper relation¹⁶ followed by Galactic and M31 globular clusters. The dotted lines show the path predicted by our ‘threshing’ scenario for nucleated dwarfs being stripped to form UCDs. The small solid symbols show the locations of dwarf galaxy nuclei after model subtraction of their parent galaxies. All the data are taken from the literature (globular clusters¹⁶, ω Centauri²⁷, ellipticals²⁸, nucleated dwarf ellipticals¹⁴) except for our measurements of the UCDs and FCC303. Note that the UCDs fill a previously empty region of the diagram between the globular clusters and galaxies. The velocity dispersions were obtained from high-resolution spectra using the VLT and the W.M. Keck Telescope echelle spectrographs. We measured velocity dispersions using standard cross-correlations of the object spectra with stellar template spectra to determine both a redshift and a correlation width. The correlation width was used to estimate the velocity dispersion by comparison with results from artificially broadening the template stars by convolution with gaussian filters of known widths.

different to both nucleated dwarf galaxies and the most luminous globular star clusters. Even if UCDs were formed from dE,N galaxies, as proposed by the threshing hypothesis, they constitute a new and distinct type of galaxy, just as spirals stripped of their star-forming abilities constitute the galaxy type S0 (ref. 21), and merging disk galaxies are believed to form ellipticals. The discovery of these objects confirms the prediction⁷ that very compact galaxies exist that were previously misclassified as foreground stars.

Although we have shown that UCDs are dynamically distinct from globular star clusters, the two can only be distinguished at the distance of the Fornax cluster by high-resolution imaging and spectroscopy. Any UCDs captured by the central elliptical may therefore be mistaken for ordinary globular star clusters in existing studies. The two populations could be separated, in principle, by observations similar to those we described here, but the requisite spectroscopy for fainter objects would be challenging with existing 8–10-m telescopes. Any dismantled dwarf elliptical galaxies captured by the central galaxy will also contribute ordinary globular star clusters to the mix. Together, these processes may explain the unusually high frequency of globular star clusters in the central elliptical galaxies of galaxy clusters²² because these dwarf galaxies also contain higher than average globular cluster populations²³.

Searches for UCDs in other rich environments, such as the populous Virgo cluster as well as dense groups, will clarify the nature and origin of UCDs while also potentially exploring the role of tidal disruption processes in galaxy and cluster evolution. □

Received 5 February; accepted 15 April 2003; doi:10.1038/nature01666.

- Moore, B., Katz, N., Lake, G., Dressler, A. & Oemler, A. Galaxy harassment and the evolution of clusters of galaxies. *Nature* **379**, 613–616 (1996).
- Bassino, L. P., Muzzio, J. C. & Rabolli, M. Are globular clusters the nuclei of cannibalized dwarf galaxies? *Astrophys. J.* **431**, 634–639 (1994).
- Bekki, K., Couch, W. J. & Drinkwater, M. J. Galaxy threshing and the formation of ultracompact dwarf galaxies. *Astrophys. J.* **552**, L105–L108 (2001).
- Gregg, M. D. & West, M. J. Galaxy disruption as the origin of intracluster light in the Coma cluster of galaxies. *Nature* **396**, 549–552 (1998).
- Durrell, P. R., Ciardullo, R., Feldmeier, J. J., Jacoby, G. H. & Sigurdsson, S. Intracuster red giant stars I the Virgo cluster. *Astrophys. J.* **570**, 119–131 (2002).
- Ford, H., Peng, E. & Freeman, K. In *ASP Conf. Ser. 273: The Dynamics, Structure and History of Galaxies: A Workshop in Honour of Professor Ken Freeman* 41 (Astronomical Society of the Pacific, San Francisco, 2002).
- Disney, M. J. Visibility of galaxies. *Nature* **263**, 573–575 (1976).
- Drinkwater, M. J. *et al.* The Fornax spectroscopic survey. I. Survey strategy and preliminary results on the redshift distribution of a complete sample of stars and galaxies. *Astron. Astrophys.* **355**, 900–914 (2000).
- Drinkwater, M. J., Jones, J. B., Gregg, M. D. & Phillipps, S. Compact stellar systems in the Fornax cluster: Super-massive star clusters or extremely compact dwarf galaxies? *Publ. Astron. Soc. Aust.* **17**, 227–233 (2000).
- Phillipps, S., Drinkwater, M. J., Gregg, M. D. & Jones, J. B. Ultracompact dwarf galaxies in the Fornax cluster. *Astrophys. J.* **560**, 201–206 (2001).
- Drinkwater, M. J., Gregg, M. D. & Colless, M. Substructure and dynamics of the Fornax cluster. *Astrophys. J.* **548**, L139–L142 (2001).
- Mieske, S., Hilker, M. & Infante, L. Ultra compact objects in the Fornax cluster of galaxies: Globular clusters of dwarf galaxies? *Astron. Astrophys.* **383**, 823–837 (2002).
- Lotz, J. M. *et al.* Dynamical friction in DE globular cluster systems. *Astrophys. J.* **552**, 572–581 (2001).
- Geha, M., Guhathakurta, P. & van der Marel, R. P. Internal dynamics, structure, and formation of dwarf elliptical galaxies. I. A Keck/Hubble Space Telescope study of six Virgo cluster dwarf galaxies. *Astron. J.* **124**, 3073–3087 (2002).
- Drinkwater, M. & Hardy, E. Extreme blue compact dwarf galaxies in the Virgo cluster. *Astron. J.* **101**, 94–101 (1991).
- Djorgovski, S. G. *et al.* Dynamical correlations for globular clusters in M31. *Astrophys. J.* **474**, L19–L22 (1997).
- Meylan, G. *et al.* Mayall II = G1 in M31: Giant globular cluster or core of a dwarf elliptical galaxy? *Astron. J.* **122**, 830–841 (2001).
- Hernquist, L. Performance characteristics of tree codes. *Astrophys. J. Suppl.* **64**, 715–734 (1987).
- Gnedin, O. Y. *et al.* The unique history of the globular cluster ω Centauri. *Astrophys. J.* **568**, L23–L26 (2002).
- Faber, S. M. & Jackson, R. E. Velocity dispersions and mass-to-light ratios for elliptical galaxies. *Astrophys. J.* **204**, 668–683 (1976).
- Bekki, K., Couch, W. J. & Shioya, Y. Passive spiral formation from halo gas starvation: Gradual transformation into S0s. *Astrophys. J.* **577**, 651–657 (2002).
- Hilker, M., Infante, L. & Richtler, T. The central region of the Fornax cluster. III. Dwarf galaxies, globular clusters, and cD halo—are there interrelations? *Astron. Astrophys. Suppl.* **138**, 55–70 (1999).
- Miller, B. W., Lotz, J. M., Ferguson, H. C., Stiavelli, M. & Whitmore, B. C. The specific globular cluster frequencies of dwarf elliptical galaxies from the Hubble Space Telescope. *Astrophys. J.* **508**, L133–L137 (1998).
- Hilker, M., Mieske, S. & Infante, L. Faint dwarf spheroidals in the Fornax cluster: A flat luminosity function. *Astron. Astrophys.* **397**, L9–L12 (2003).
- Ferguson, H. C. Population studies in groups and clusters of galaxies. II—A catalog of galaxies in the central 3.5 deg of the Fornax cluster. *Astron. J.* **98**, 367–418 (1989).
- Larsen, S. S. Young massive star clusters in nearby galaxies. II. Software tools, data reductions and cluster sizes. *Astron. Astrophys. Suppl.* **139**, 393–415 (1999).
- Merritt, D., Meylan, G. & Mayor, M. The stellar dynamics of omega centauri. *Astron. J.* **114**, 1074–1086 (1997).
- Burstein, D., Bender, R., Faber, S. & Nolthenius, R. Global relationships among the physical properties of stellar systems. *Astron. J.* **114**, 1365–1392 (1997).

Acknowledgements This paper is based on observations made with the HST, the European Southern Observatory VLT, the W.M. Keck Telescope, and the Las Campanas Observatory 100-inch du Pont Telescope. This work is supported by the Australian Research Council and the Australian Nuclear Science and Technology Organisation. M.D.G. acknowledges support from the National Science Foundation and from the Space Telescope Science Institute and NASA which is operated by AURA, Inc. Part of this work was performed under the auspices of the US Department of Energy by University of California Lawrence Livermore National Laboratory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.D. (mjd@physics.uq.edu.au).

Antiferromagnetic order as the competing ground state in electron-doped Nd_{1.85}Ce_{0.15}CuO₄

H. J. Kang*, Pengcheng Dai†, J. W. Lynn‡, M. Matsuura†, J. R. Thompson*, Shou-Cheng Zhang§, D. N. Argyriou||, Y. Onose¶ & Y. Tokura¶#☆

* Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996-1200, USA

† Condensed Matter Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6393, USA

‡ NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8562, USA

§ Department of Physics, McCullough Building, Stanford University, Stanford, California 94305-4045, USA

|| Hahn-Meitner-Institut, Glienicker Str 100, Berlin D-14109, Germany

¶ Spin Superstructure Project, ERATO, Japan Science and Technology, Tsukuba 305-8562, Japan

Correlated Electron Research Center, Tsukuba 305-8562, Japan

☆ Department of Applied Physics, University of Tokyo, Tokyo 13-8656, Japan

Superconductivity in the high-transition-temperature (high- T_c) copper oxides competes with other possible ground states^{1,2}. The physical explanation for superconductivity can be constrained by determining the nature of the closest competing ground state, and establishing if that state is universal among the high- T_c materials. Antiferromagnetism has been theoretically predicted^{3,4} to be the competing ground state. A competing ground state is revealed when superconductivity is destroyed by the application of a magnetic field, and antiferromagnetism has been observed in hole-doped materials under the influence of modest fields^{5–12}. None of the previous experiments have revealed the quantum phase transition from the superconducting state to the antiferromagnetic state, because they failed to reach the upper critical field B_{c2} . Here we report the results of transport and neutron-scattering experiments on electron-doped Nd_{1.85}Ce_{0.15}CuO₄ (refs 13, 14), where B_{c2} can be reached¹⁵. The applied field reveals a static, commensurate, anomalously conducting long-range ordered antiferromagnetic state, in which the induced moment scales approximately linearly with the field strength until it saturates at B_{c2} . This and previous experiments on the hole-doped materials therefore establishes antiferromagnetic order as a competing ground state in the high- T_c copper oxide materials, irrespective of electron or hole doping.

We grew single crystals of Nd_{1.85}Ce_{0.15}CuO₄ (NCCO) using the travelling solvent floating-zone technique¹⁶. As grown, the samples are not superconducting^{17,18} and exhibit the non-collinear long-range antiferromagnetic (AF) structure of undoped Nd₂CuO₄ (Fig. 1a; refs 19–21). To obtain superconductivity, the samples were annealed in an Ar/O₂ gas mixture¹⁶. Although antiferromagnetism in the heat-treated NCCO is greatly reduced, it could not be completely eliminated, and the residual AF order coexists with superconductivity²². Figure 1c, d shows the temperature (T) dependence of the in-plane resistivity (ρ_{ab}) at zero field and the magnetic susceptibility of the sample ($T_c = 25$ K with a transition width of 3 K), respectively. At 5 K, $\rho_c(B)$ measured for the c -axis-aligned field ($B||c$ -axis) shows a field-induced superconducting-to-normal state transition at $B_{c2} = 6.2$ T (Fig. 1e). Figure 1f displays $\rho_c(T)$ for several applied fields. In a field (9 T) that exceeds B_{c2} , $\rho_c(T)$ initially decreases as temperature decreases but shows an anomalous upturn for $T < 12$ K, similar to Pr_{1.85}Ce_{0.15}CuO₄ (refs 23, 24).

We use neutron diffraction to investigate the magnetic ordering, and label wave vectors $Q = (q_x, q_y, q_z)$ in Å⁻¹ as $(H, K, L) = (q_x a/2\pi, q_y a/2\pi, q_z c/2\pi)$ in the reciprocal lattice units (r.l.u.) appropriate for

the tetragonal unit cell of NCCO (space group $I4/mmm$, $a = 3.92$ Å and $c = 12.07$ Å). The experiments were performed in two different geometries. First, the superconducting CuO₂ planes were aligned in the horizontal ($H, K, 0$) scattering plane and the applied vertical field was along the c -axis ($B||c$ -axis). Second, the crystal was aligned in the $[H, H, L]$ zone and the vertical field was applied along the $[1, -1, 0]$ direction ($B||ab$ -plane). In common with other layered superconductors, a $B||c$ -axis field suppresses superconductivity more strongly than a $B||ab$ -plane field.

Figures 1g, h, 2a–f and 3a–h summarize the outcome of the $B||c$ -axis experiments. At 55 K, we observe weak structural superlattice peaks intrinsic to superconducting NCCO (Fig. 2a–c; ref. 25). While long-range ferromagnetic (FM) ordering is induced—as seen by the added magnetic intensity to the $(1, 1, 0)$ structural Bragg peak intensity—when a 6-T $B||c$ -axis field is applied (Fig. 2d), the lack of intensity change at superlattice positions (Fig. 2a–c) shows that no static AF order is induced by the field at this temperature. On cooling to 5 K, the Cu spins order below a Néel temperature T_N of ~ 38 K in the non-collinear structure (Figs 1a and 2b; refs 19–21). The rapid intensity increase below 15 K in Fig. 3b at 0 T is due to the polarization of the Nd moment by strong Cu–Nd interactions^{17–19}. At 5 K and 6 T, in addition to an enhanced FM moment at $(1, 1, 0)$ (Fig. 2d), field-induced signals begin to emerge at superlattice positions (Fig. 2a–c). Particularly interesting is the appearance of magnetic peaks at $(0.5, 0.5, 0)$ and $(0.5, 0, 0)$, which demonstrates that the field-induced magnetic structure differs from that in zero field (Fig. 1a), as the latter has vanishing intensity at these positions (Fig. 1g; ref. 19). A survey of reciprocal space in the $(H, K, 0)$ plane indicates clear field-induced effects that obey the selection rules $(\pm(2m+1)/2, \pm(2n+1)/2, 0)$, $(\pm(2m+1)/2, \pm n, 0)$, and $(\pm m, \pm(2n+1)/2, 0)$ with $m, n = 0, 1, 2$ (Fig. 1h). Table 1 summarizes these observations, and makes intensity comparisons with model calculations.

For AF ordered copper oxides, the magnetic scattering may have zero or non-zero intensity at $(0.5, 0.5, 0)$ depending on the spin arrangements along the c -axis, but it always vanishes at $(0.5, 0, 0)$. To understand the anomalous field-induced $(0.5, 0, 0)$ scattering, we consider four different models. First, the sample may phase-separate into two different magnetic domains with one domain contributing to $(0.5, 0.5, 0)$ and the other to $(0.5, 0, 0)$. Second, two CuO₂ sheets within the unit cell may have different magnetic structures, with one contributing to $(0.5, 0.5, 0)$ and the other to $(0.5, 0, 0)$. However, it is unclear why two identical CuO₂ planes should behave differently in an external field. Third, assuming that structural superlattice reflections (Table 1) originate from lattice distortions of both Cu (Nd) and O (ref. 25), the field-induced effect at these positions may simply be FM ordering associated with the polarization of Cu(Nd) ions. The differences in the field-dependent data between superlattice (Fig. 3e–g) and $(1, 1, 0)$ Bragg (Fig. 3h) reflections, though, are inconsistent with this model. Fourth, the Cu

Table 1 Experimentally observed integrated intensities and model calculations

(H, K, L)	I_{lattice} at 55 K	I_{mag} at 5 K	I_{mag} (model 4)*
(0.5, 0.5, 0)	1.7	7.1	7.1
(1.5, 1.5, 0)	1.8	0.3	1.7
(0, 0.5, 0)	0.5	7.8	6.92
(0, 1.5, 0)	< 0.16	5.5	1.9
(1, 0.5, 0)	0.4	2.1	2.8
(0.5, 1, 0)	0.4	3.2	2.8
(2, 0.5, 0)	0.3	0.5	1.2
(0.5, 2, 0)	< 0.16	1.4	1.2
(1, 1.5, 0)	0.25	1.1	1.5
(1.5, 1, 0)	< 0.16	0.8	1.5
(0, 1, 0)	1.1	< 0.16	0

Extinction and Nd absorption for the large crystal (a cylindrical rod 38 mm long and 3.8 mm in diameter) caused intensity (in arbitrary units) differences for equivalent reflections. The uncertainties for the intensities were about 0.16.

* See text.

moments in NCCO could be different at different sites within the CuO_2 plane, partially owing to the intrinsically inhomogeneous doping by electrons. Assuming three different Cu moments M_1 , M_2 and M_3 in the type-II AF structure¹⁹, we find that the spin arrangements in Fig. 1b can best describe the observed scattering (Table 1). Although this model may not uniquely describe the data, our results should stimulate future work in this direction.

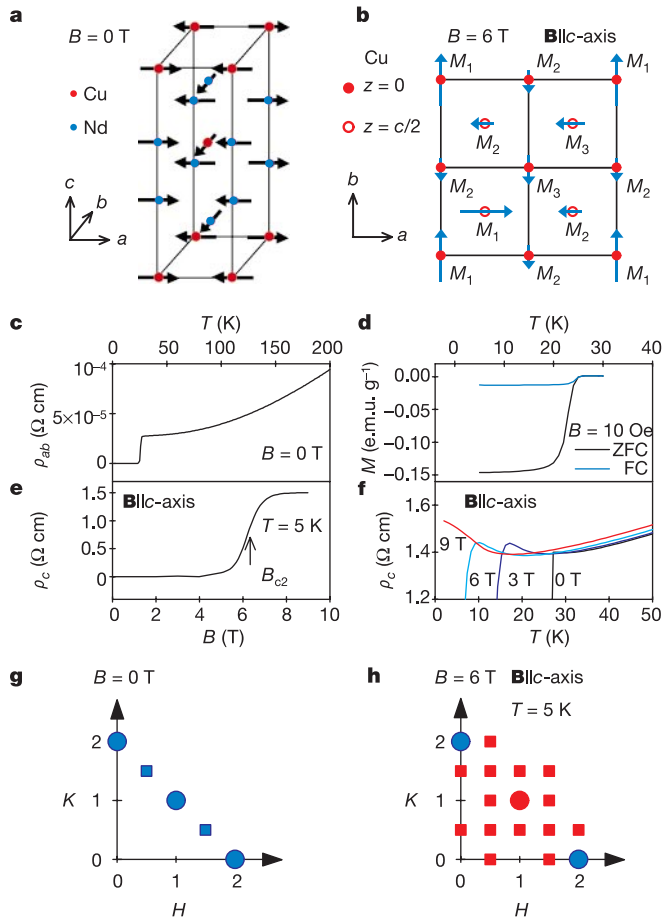


Figure 1 Spin structure models, bulk transport (magnetic) properties, and summary of reciprocal space probed in the neutron experiments. For resistivity and magnetic susceptibility, we used a 9-T physical property measurement system and a 7-T SQUID magnetometer, respectively. Our neutron-scattering experiments were performed on the BT-9 and BT-2 triple-axis spectrometers at the NIST Center for Neutron Research and on the E4 two-axis diffractometer at the Berlin Neutron Scattering Center, Hahn-Meitner-Institute (HMI). For BT-9, the collimations were, proceeding from the reactor to the detector, 40° - 46° - 40° - 80° (full-width at half-maximum), and the final neutron energy was fixed at $E_f = 14.78$ meV. Similar collimations were used for BT-2 measurements. The monochromator, analyser and filters were all pyrolytic graphite. The low- T data were collected after field cooling the sample from 15–25 K above T_c . **a**, The non-collinear type-II AF spin structure of the AF order in NCCO below 39 K (ref. 19). **b**, Model 4 (see text). For clarity of the AF spin arrangements, we ignored the field-induced FM moments along the c -axis. As we only collected data in the $(H, K, 0)$ plane for $\mathbf{B} \parallel c$ -axis, no information was obtained for the spin arrangements along the c -axis direction. **c**, $\rho_{ab}(T)$ at 0 T. **d**, T dependence of the zero-field-cooled (ZFC) and field-cooled (FC) magnetic susceptibility in 10-Oe field. **e**, Magnetic field dependence of ρ_c at 5 K. **f**, $\rho_c(T)$ at 0, 3, 6 and 9 T. **g**, $\rho_{ab}(T)$ behaves similarly. The blue circles and squares in **g** show the observed fundamental nuclear and magnetic peaks, respectively, at 0 T. In the c -axis spin arrangements of **a**, magnetic structure factors have non-zero values only at $(\pm 0.5, \pm 1.5, 0)$ and $(\pm 1.5, \pm 0.5, 0)$ for $L = 0$. For clarity, we did not plot the weak nuclear superlattice reflections. The red squares and circles in **h** illustrate the field-induced new magnetic peaks or intensity gain at 6 T.

Figure 3 summarizes the temperature and field dependence of the neutron intensity at $(0.5, 0.5, 0)$, $(0.5, 1.5, 0)$, $(0.5, 0, 0)$ superlattice and $(1, 1, 0)$ structural Bragg peak positions. While the field-induced AF order peaks around 6.5 T at $(0.5, 0.5, 0)$ indicative of a phase transition from the superconducting state to an antiferromagnetically ordered state at B_{c2} (Figs 2e and 3e)—the field-induced FM intensity at $(1, 1, 0)$ continues to rise for fields above B_{c2} , and does not saturate for fields up to $2B_{c2}$ (Figs 2f, 3d and 3h).

We believe that the field-induced AF signal arises from the suppression of superconductivity by the c -axis-aligned field. For Nd_2CuO_4 , a $\mathbf{B} \parallel ab$ -plane field induces a spin-flop transition from the non-collinear to a collinear structure²⁰. At 0 T and 5 K, magnetic scattering for the spin structure of Fig. 1a has vanishing intensity at $(0.5, 0.5, 0)$ and becomes very strong at $(0.5, 0.5, 3)$ (ref. 19). The observed L dependence of the data is consistent with this spin structure (Fig. 4a–d). On applying a 7-T $\mathbf{B} \parallel ab$ -plane field, the zero-field intensities for $(0.5, 0.5, 1)$ and $(0.5, 0.5, 3)$ disappear. This is expected when the residual non-collinear AF order in NCCO is

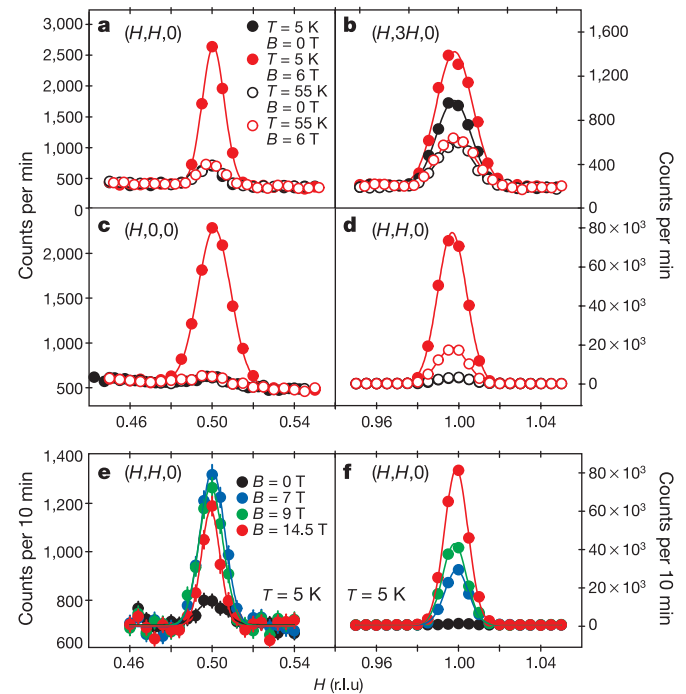


Figure 2 Effect of a $\mathbf{B} \parallel c$ -axis field on the magnetic peaks (half integer) and induced ferromagnetic (integer) peaks below and above T_c . The data in **a–d** were collected at NIST, whereas the high-field data in **e** and **f** were obtained on the E4 diffractometer using the VM-1, 14.5-T vertical field magnet at HMI. The collimations were 40° - 40° -sample- 40° with one pyrolytic graphite filter. Scan directions are along **a**, **d**, $(H, H, 0)$; **b**, $(H, 3H, 0)$; **c**, $(H, 0, 0)$; **e**, **f**, $(H, H, 0)$. In **a–d**, the black and red filled circles represent data at 5 K for identical scans at zero field and 6-T field, respectively. The black and red open circles represent data at 55 K for identical scans at zero field and 6-T field, respectively. The Q-widths of the zero-field and field-induced scattering are resolution-limited and identical, implying an in-plane correlation length larger than 300 Å. The peaks in open black circles of **a–c** are due entirely to structural superlattice scattering. At zero field, we confirmed the weak structural superlattice reflections reported in ref. 25 at $(H, K, L) = (\pm(2m+1)/2, \pm(2n+1)/2, 0)$ where $m, n = 0, 1$. We also observed superlattice reflections at $(\pm 0.5, 0, 0)$ type positions not allowed in the $I4/mmm$ space group of NCCO. Along the c -axis (L) direction, these superlattice peaks are very broad. At present, the microscopic origin of these superlattice reflections is unknown²⁵. **e**, $(H, H, 0)$ scans through $(0.5, 0.5, 0)$ at 5 K as a function of field for fields up to 14.5 T. The field-induced moment increases with increasing field up to B_{c2} . For fields above B_{c2} , the field-induced signal begins to drop due to the canting of field-induced antiferromagnetism. **f**, $(H, H, 0)$ scans through $(1, 1, 0)$ at 5 K show that the field-induced FM moment does not saturate up to 14.5 T.

transformed into a collinear spin structure by the applied field. For Nd_2CuO_4 , the critical field for such a spin-flop transition is 0.8 T at 5.6 K (ref. 20). This is consistent with the spin-flop critical field shown in Fig. 4f. Finally, Fig. 4e shows the T dependence of the scattering at (0.5, 0.5, 3) in zero field and at 7 T. At 7 T, the intensity shows no T dependence below 65 K, indicating that this scattering is non-magnetic. Comparison of zero-field and 7-T data gives $T_N \approx 38$ K at 0 T, a value larger than T_c but substantially smaller than that of other NCCO samples^{17,18}. As such a field has a limited effect on superconductivity and does not change the spin moments (Fig. 4), we conclude that the field-induced AF structure seen in the

B|| c -axis geometry is a direct consequence of the suppression of superconductivity.

Understanding the quantum phase transition from the AF parent state to the superconducting state is the key issue for high- T_c copper oxides. Traditionally, this transition is achieved by chemical doping, which introduces extrinsic effects that mask the nature of the transition. Magnetic field provides an alternative, much cleaner, tuning parameter in high- T_c superconductors^{1,2}. Although previous experiments reported enhanced AF fluctuations in the vortex state^{5–12}, there is always a logical possibility that other forms of order are induced at higher fields, and compete predominantly with

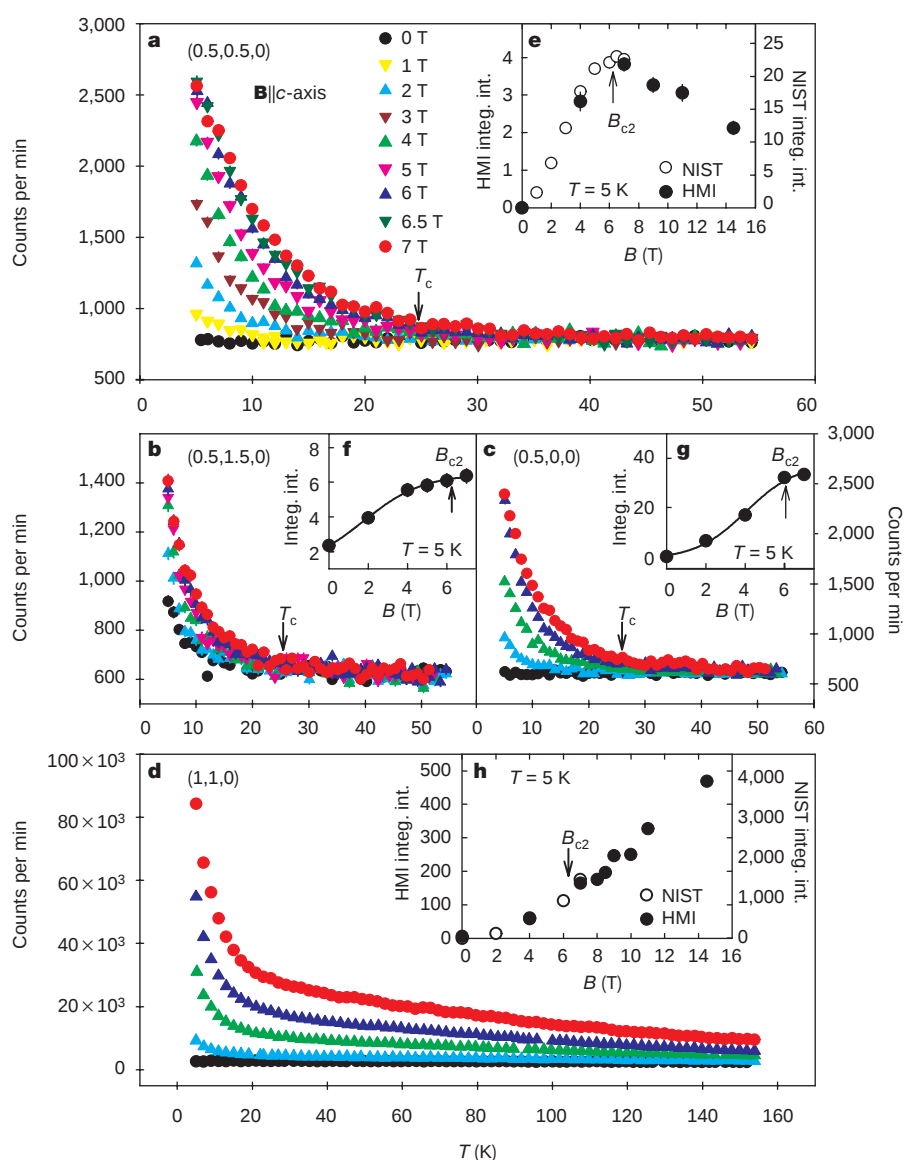


Figure 3 Effect of a **B**|| c -axis field on the T dependence of the scattering at various positions and the field dependence of the integrated intensity at 5 K. **a–d**, T dependence of the scattering at (0.5, 0.5, 0), (0.5, 1.5, 0), (0.5, 0, 0) and (1, 1, 0). At 0 T, the intensity at (0.5, 0.5, 0) in **a** is essentially temperature independent below 55 K. On application of a field, the scattering begins to increase systematically above the zero-field value from $T \leq 8$ K at 1 T to $T \leq 25$ K at 6 T, and finally saturates for fields around 6.5 T. **e**, The field-induced integrated magnetic intensity of the (0.5, 0.5, 0) peak as a function of applied field. The open circles are data taken at NIST whereas filled circles are results from HMI. These two sets of data are completely consistent below 7 T. For $B > B_{c2}$, the

field-induced signal decreases with increasing field due to the canting of field-induced AF moment. **f, g**, The field-induced scattering at (0.5, 1.5, 0) and (0.5, 0, 0). **h**, The field-induced FM moment at (1, 1, 0). The field-induced FM moment increases from 0 T to 14.5 T, and shows no major anomaly around B_{c2} . By normalizing the T -dependent intensity of (0.5, 0.5, 0) and (0.5, 0, 0) at 6 T with that of the (1, 1, 0) peak at zero field, we estimate Cu moments (in μ_B) of $M_1 \approx 0.19$, $M_3 \approx 0.07$ and $M_2 \approx 0.05$ at 15 K, and Nd moments (in μ_B) of $M_{1\text{Nd}} \approx 0.13$, $M_{3\text{Nd}} \approx 0.053$ and $M_{2\text{Nd}} \approx 0.037$ at 5 K, using the spin arrangements given in Fig. 1b. The FM field-induced moment increases with field approximately linearly up to a field close to 14.5 T.

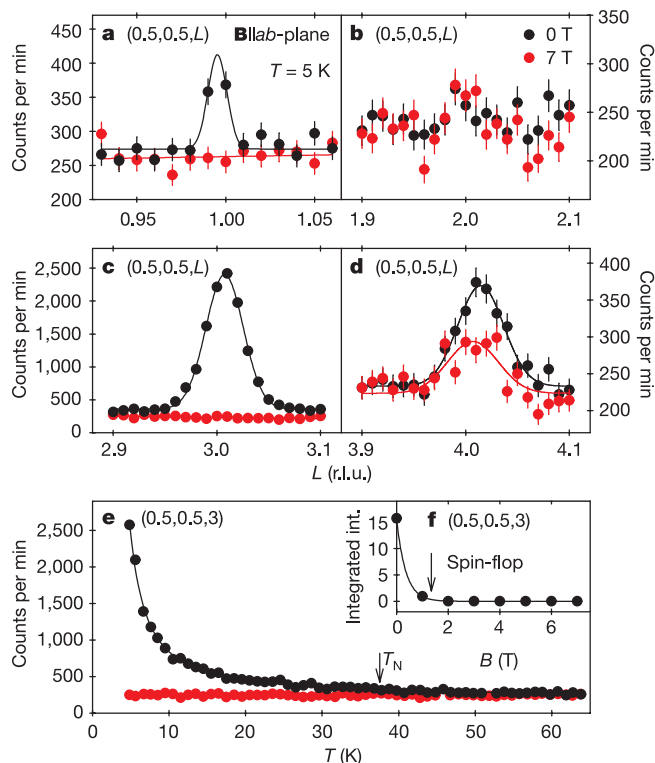


Figure 4 Effect of a $B||ab$ -plane field on $(0.5, 0.5, L)$ reflections. At 0 T and 5 K, scans along the $[H, H, 0]$ direction show resolution-limited structural superlattice peaks around $(0.5, 0.5, 0)$ (Fig. 2b) and $(1.5, 1.5, 0)$. Along the c -axis (L) direction, these peaks are very broad²⁵. **a–d**, Scans along the $[0.5, 0.5, L]$ direction at zero (black circles) and 7 T (red circles). On applying a 7-T $B||ab$ -plane field, the zero-field peak at $(0.5, 0.5, 3)$ disappears because of the spin-flop transition. We confirmed that a 4-T c -axis-aligned field enhances the $(0.5, 0.5, 3)$ reflection (unpublished data taken on the E4 diffractometer using the HM2 4-T horizontal field magnet at HMI). **e**, T dependence of the scattering at $(0.5, 0.5, 3)$ at zero field and 7 T with T_N (~ 38 K) marked by the arrow. **f**, The field dependence of the integrated intensity at $(0.5, 0.5, 3)$. The critical field for the spin-flop transition is marked by the arrow. At zero field, $M_{Cu} \approx 0.016 \mu_B$ and $M_{Nd} \approx 0.01 \mu_B$ were estimated by normalizing the integrated intensities of $(0.5, 0.5, 1)$ and $(0.5, 0.5, 3)$ to that of $(1, 1, 0)$ (ref. 22). At 7 T, we obtained $M_{Cu} \approx 0.012 \mu_B$ and $M_{Nd} \approx 0.01 \mu_B$ by using $(0.5, 0.5, 2)$, $(0.5, 0.5, 4)$ and $(1, 1, 0)$ reflections. Therefore, the spin-flop transition induced by the 7-T in-plane field does not change the spin moments.

superconductivity at B_{c2} . Our experiment demonstrates that a direct quantum phase transition from the superconducting state to an anomalously conducting antiferromagnetically ordered state is induced at B_{c2} . Combined with earlier reports on the hole-doped materials^{5–12}, this experiment shows that AF order can be induced by an applied field irrespective of electron or hole doping. The detailed experimental results reported here, including the temperature and field dependence of the induced AF moments, provide powerful quantitative constraints on our theoretical understanding of the interplay between magnetism and superconductivity in high- T_c copper oxides. □

Received 11 November 2002; accepted 1 April 2003; doi:10.1038/nature01641.

- Levi, B. G. Magnetism and superconductivity fight for control in high- T_c superconductors. *Phys. Today* **55**, 14–15 (2002).
- Sachdev, S. & Zhang, S. C. Tuning order in the cuprate superconductors by a magnetic field. *Science* **295**, 452–454 (2002).
- Zhang, S. C. A unified theory based on SO(5) symmetry of superconductivity and antiferromagnetism. *Science* **275**, 1089–1096 (1997).
- Arovas, D. P., Berlinsky, A. J., Kallin, C. & Zhang, S. C. Superconducting vortex with antiferromagnetic core. *Phys. Rev. Lett.* **79**, 2871–2874 (1997).
- Katano, S., Sato, M., Yamada, K., Suzuki, T. & Fukase, T. Enhancement of static antiferromagnetic correlations by magnetic field in a superconductor $La_{2-x}Sr_xCuO_4$ with $x = 0.12$. *Phys. Rev. B* **62**, R14677–R14680 (2000).

- Lake, B. *et al.* Spins in the vortices of a high-temperature superconductor. *Science* **291**, 1759–1762 (2001).
- Lake, B. *et al.* Antiferromagnetic order induced by an applied magnetic field in a high-temperature superconductor. *Nature* **415**, 299–302 (2002).
- Khaykovich, B. *et al.* Enhancement of long-range magnetic order by magnetic field in superconducting La_2CuO_{4+y} . *Phys. Rev. B* **66**, 014528 (2002).
- Miller, R. I. *et al.* Evidence for static magnetism in the vortex cores of ortho-II $YBa_2Cu_3O_{6.50}$. *Phys. Rev. Lett.* **88**, 137002 (2002).
- Mitrovic, V. F. *et al.* Spatially resolved electronic structure inside and outside the vortex cores of a high-temperature superconductor. *Nature* **413**, 501–504 (2001).
- Mitrovic, V. F. *et al.* Antiferromagnetism in the vortex cores of $YBa_2Cu_3O_{7-\delta}$. Preprint cond-mat/0202368 at (<http://xxx.lanl.gov>) (2002).
- Kakuyanagi, K., Kumagai, K., Matsuda, Y. & Hasegawa, M. Antiferromagnetic ordering and disappearance of pseudogap within the vortex core of $Tl_2Ba_2CuO_{6+\delta}$. Preprint cond-mat/0206362 at (<http://xxx.lanl.gov>) (2002).
- Tokura, Y., Takagi, H. & Uchida, S. A superconducting copper oxide compound with electrons as the charge carriers. *Nature* **337**, 345–347 (1989).
- Takagi, H., Uchida, S. & Tokura, Y. Superconductivity produced by electron doping in CuO_2 -layered compounds. *Phys. Rev. Lett.* **62**, 1197–1200 (1989).
- Hidaka, Y. & Suzuki, M. Growth and anisotropic superconducting properties of $Nd_{2-x}Ce_xCuO_{4-y}$ single crystals. *Nature* **338**, 635–637 (1989).
- Onose, Y., Taguchi, Y., Ishizaka, K. & Tokura, Y. Doping dependence of pseudogap and related charge dynamics in $Nd_{2-x}Ce_xCuO_4$. *Phys. Rev. Lett.* **87**, 217001 (2001).
- Yamada, K. *et al.* Neutron scattering study on electron-hole doping symmetry of high- T_c superconductivity. *J. Phys. Chem. Solids* **60**, 1025–1030 (1999).
- Uefuji, T. *et al.* Coexistence of antiferromagnetic ordering and high- T_c superconductivity in electron-doped superconductor $Nd_{2-x}Ce_xCuO_4$. *Physica C* **357**, 208–211 (2001).
- Lynn, J. W. & Skanthakumar, S. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 31 (eds Gschneidner, K. A. Jr, Eyring, L. & Maple, M. B.) 315–350 (Elsevier Science, Amsterdam, 2001).
- Skanthakumar, S., Lynn, J. W., Peng, J. L. & Li, Z. Y. Field-dependence of the magnetic-ordering of Cu in R_2CuO_4 ($R = Nd, Sm$). *J. Appl. Phys.* **73**, 6326–6328 (1993).
- Skanthakumar, S., Lynn, J. W., Peng, J. L. & Li, Z. Y. Observation of noncollinear magnetic-structure for the Cu spins in Nd_2CuO_4 -type systems. *Phys. Rev. B* **47**, 6173–6176 (1993).
- Uefuji, T., Kurahashi, K., Fujita, M., Matsuda, M. & Yamada, K. Electron-doping effect on magnetic order and superconductivity in $Nd_{2-x}Ce_xCuO_4$ single crystals. *Physica C* **378–381**, 273–277 (2002).
- Fournier, P. *et al.* Insulator-metal crossover near optimal doping in $Pr_{2-x}Ce_xCuO_4$: Anomalous normal-state low temperature resistivity. *Phys. Rev. Lett.* **81**, 4720–4723 (1998).
- Hill, R. W. *et al.* Breakdown of Fermi-liquid theory in a copper-oxide superconductor. *Nature* **414**, 711–715 (2001).
- Kurahashi, K., Matsushita, H., Fujita, M. & Yamada, K. Heat treatment effects on the superconductivity and crystal structure of $Nd_{1.85}Ce_{0.15}CuO_4$ studied using a single crystal. *J. Phys. Soc. Jpn* **71**, 910–915 (2002).

Acknowledgements We thank R. Jin, D.-H. Lee, S.-H. Lee, H. A. Mook and S. Skanthakumar for discussions, and R. Jin and K. Prokes for experimental assistance. This work was supported by the US NSF and DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D. (daip@ornl.gov).

Selectivity in vibrationally mediated single-molecule chemistry

J. I. Pascual^{*†}, N. Lorente[‡], Z. Song^{*§}, H. Conrad^{*} & H.-P. Rust^{*}

^{*}Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14194 Berlin, Germany

[†]Institut de Ciència de Materials de Barcelona-CSIC, Campus de la UAB, E-08193 Bellaterra, Spain

[‡]Laboratoire Collisions, Agrégats, Réactivité, UMR5589, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France

[§]Department of Chemistry, Brookhaven National Laboratory, Building 555, PO Box 5000, Upton, New York 11973-5000, USA

The selective excitation of molecular vibrations provides a means to directly influence the speed and outcome of chemical reactions. Such mode-selective chemistry¹ has traditionally used laser pulses to prepare reactants in specific vibrational states² to enhance reactivity^{3,4} or modify the distribution of product species^{5,6}. Inelastic tunnelling electrons may also excite molecular vibrations^{7,8} and have been used to that effect on adsorbed

molecules, to cleave individual chemical bonds⁹ and induce molecular motion^{10–13} or dissociation^{14–17}. Here we demonstrate that inelastic tunnelling electrons can be tuned to induce selectively either the translation or desorption of individual ammonia molecules on a Cu(100) surface. We are able to select a particular reaction pathway by adjusting the electronic tunnelling current and energy during the reaction induction such that we activate either the stretching vibration of ammonia or the inversion of its pyramidal structure. Our results illustrate the ability of the scanning tunnelling microscope to probe single-molecule events in the limit of very low yield and very low power irradiation, which should allow the investigation of reaction pathways not readily amenable to study by more conventional approaches.

Single NH₃ molecules adsorbed on a Cu(100) surface (Fig. 1a) sit on top of a copper atom with the three-fold symmetry axis oriented perpendicular to the surface. Interaction between the nitrogen lone-pair electrons and the metal surface forms a chemical bond, which is strong enough for stable imaging by a scanning tunnelling microscope (STM) at the low temperature of our instrument¹⁸ (5 K). However, controlled cleavage of the NH₃–Cu(100) bond can be easily induced by increasing the energy of the tunnelling electrons (eV_s , where $-e$ is the charge on an electron and V_s is the sample bias) above well-defined threshold values. The result of this unimolecular reaction can be either the translation of the ammonia molecule onto a different adsorption site, or its desorption, as ascertained by subsequent imaging of the surrounding area (Fig. 1b).

The analysis of the electron energy thresholds connected with the induced motion of the molecule reveals the operation of two bond-breaking mechanisms, each of them activated by the excitation of a different internal vibration. In the regime of tunnelling current (I_t) values smaller than ~ 0.5 nA, the reaction mechanism (M1) presents a sharp threshold at $eV_s \approx \pm 400$ meV (Fig. 2a), which we associate with the excitation energy of the N–H stretch mode, $\nu_s(\text{N–H})$, by inelastic tunnelling electrons—the $\nu_s(\text{N–H})$ frequency is 408 meV. This assignment is confirmed by the isotopic shift of the threshold down to about ± 300 meV on the deuterated molecule, ND₃, in agreement with the frequency shift of the $\nu(\text{N–D})$ modes (Fig. 2b). The second mechanism (M2) of inducing molecular motion involves the excitation of the NH₃ umbrella or inversion mode, $\delta_s(\text{N–H}_3)$, with frequency 139 meV. The activation of this mode is deduced from the onset of an additional energy threshold at $eV_s \approx \pm 270$ meV found when we use higher tunnelling currents to induce the bond breaking (Fig. 2c). There is no fundamental vibrational state in this range of energy; instead, only a direct (one-

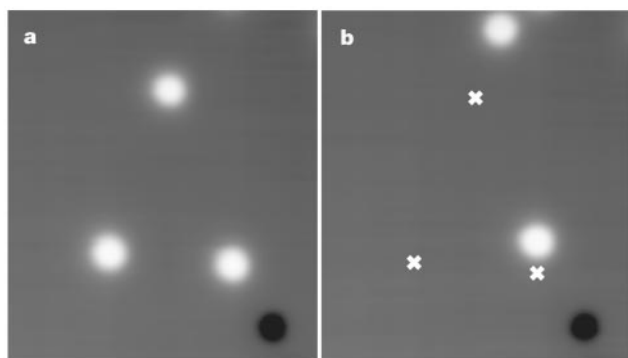


Figure 1 Inducing molecular motion with tunnelling electrons. **a**, STM image of three NH₃ molecules adsorbed on a clean Cu(100) surface. **b**, STM image of the same area after injecting 420-meV electrons ($I_t \approx 0.5$ nA) on top of each molecule. Two molecules were moved to a new site, and one was desorbed (crosses point to the initial positions of molecules in **a**). A defect in the surface (black circle at the bottom-right corner) is used as a reference point. (Conditions: $V_s = 100$ mV; $I_t = 1$ nA; field of view, 7 nm $\times$ 7 nm.)

step) excitation of the first $\delta_s(\text{N–H}_3)$ overtone can account for this value. In line with this assignment, we also find a high-current threshold for the deuterated species at the expected energy of the first $\delta_s(\text{N–D}_3)$ overtone (Fig. 2d).

Previous ultraviolet and infrared photochemistry experiments reported a mechanism of NH₃ desorption mediated by excitation of the umbrella mode, and effected by the inversion of the NH₃ molecule^{19,20}. In resemblance to this process, the ‘reaction product’ found after operation of the M2 mechanism is predominantly molecular desorption, whereas the M1 mechanism mainly results in lateral diffusion. Thus, it seems that the vibration excited by the inelastic electrons is correlated with the branching ratio between the two products found after reaction. To establish the pathway steering the vibrational energy for each mechanism, we analyse the behaviour of their reaction yield (Y_R , the probability of reaction per tunnelling electron) as a function of the tunnelling current, I_t . For a pathway overcoming a barrier with a sequence of n vibrational excitations in a ladder-climbing fashion^{21,22}, Y_R increases with the excitation rate (that is, I_t), following a power law $Y_R \propto I_t^{n-1}$. Thus, the determination of n provides experimental insight into the activation barriers for each mechanism.

Reaction mechanism M1 was selected by injecting electrons with constant energy above the $\nu_s(\text{N–H})$ mode frequency (Fig. 3a,b). The dependence of yield on I_t is shown in Fig. 3c. Y_R remains essentially constant with I_t ($n - 1 = 0$) for tunnelling current values below 0.5 nA, indicating that the M1 mechanism uses a single $\nu_s(\text{N–H})$ excitation. This process supplies an insufficient amount of vibrational energy to produce molecular desorption. Above $I_t \approx 0.5$ nA, a crossover to a two-electron process (M3) takes place. The activation of this additional mechanism corresponds to $\nu_s(\text{N–H})$ excitations, as the threshold associated with this mode in Fig. 2 persists in the regime of higher tunnelling currents. To analyse the umbrella-mediated pathway, we inject electrons with energy

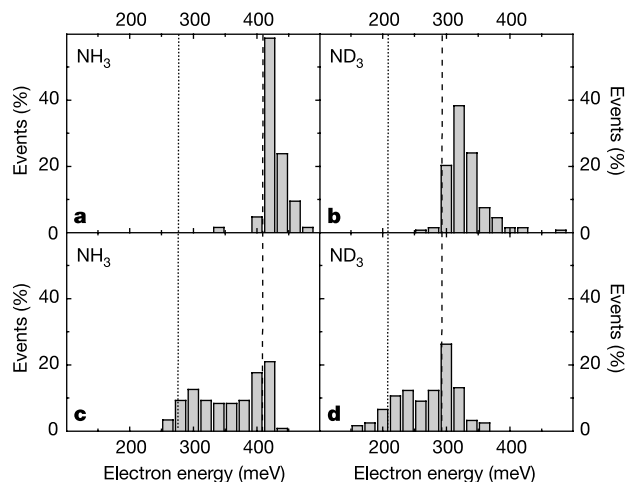


Figure 2 Thresholds of electron energy needed to induce molecular motion. The thresholds are statistically probed by fixing the STM tip above a molecule with $V_s = \pm 100$ mV, and slowly increasing the bias magnitude. A sudden current drop reveals the occurrence of a reaction. The panels show the distribution of electron energy (eV_s) reached when the reaction occurs. At low tunnelling currents ($I_t < 0.5$ nA) NH₃ shows a threshold at ~ 400 meV (**a**), which shifts down to ~ 300 meV in ND₃ (**b**), matching correspondingly the energy of N–H and N–D stretch modes (dashed lines). In this reaction mechanism (M1) 60% of the events produced molecular translation. **c**, For $I_t > 1$ nA, an additional threshold appears gradually at ~ 270 meV in NH₃, consistent with the energy of two umbrella-mode quanta (dotted lines). This reveals a second bond-cleavage mechanism (M2), which produced molecular desorption in 75% of the events. In ND₃, the corresponding onset at ~ 200 meV is observed (**d**) only after reaching tunnelling currents higher than 10 nA.

below the $\nu_s(\text{N-H})$ threshold. From the data in Fig. 3c we deduce that pathway M2 comprises a sequence of three inelastic electrons. Here, although each inelastic electron provides less energy, after three excitations the ammonia molecule accumulates sufficient energy to overcome the $\sim 600\text{-meV}$ desorption barrier.

The data shown in Fig. 3c reveal that the three mechanisms of Cu-NH₃ bond cleavage can be selected by appropriately tuning electron energy and current. In order to establish a detailed correlation between the vibrational pathways and the two types of products found, each reaction mechanism was independently selected for an ensemble of adsorbed molecules by applying the

appropriate eV_s and I_t values. The results (Fig. 3d) confirm that pathway M1 predominantly induces molecular translation, reaching a probability of $P_{\text{TRANS}} = 0.75$ for very low currents. The increasing importance of the M3 mechanism leads to a continuous change in the resulting product, until molecular desorption dominates the reaction. At the crossover ($I_t \approx 0.5\text{ nA}$), both desorption and translation are produced with similar probabilities. In contrast, desorption prevails ($P_{\text{DES}} \approx 0.8$) for the umbrella-mode-mediated pathway, M2, independently of the tunnelling current.

The origin of the mode-selective manipulation of the ammonia molecule can be rationalized by studying the activation barriers for molecular motion. Our calculations (see Methods) suggest that the lowest barrier for molecular translation is 301 meV . A single excitation of the $\nu_s(\text{N-H})$ mode is thus sufficient to overcome this barrier. Although the probability of exciting this mode with tunnelling electrons is low (6×10^{-4}), it is still much larger than the yield measured for pathway M1. However, as this vibration does not produce molecular motion along the translational coordinates, an additional factor must be included to account for the probability of transferring the $\nu_s(\text{N-H})$ vibrational energy into translational mode levels above the 301 meV barrier (Fig. 4a). Previous STM experiments^{11,12} found that internal energy transfer via anharmonic coupling between two modes can be an important decay channel for an internal vibration to induce motion of an adsorbed molecule. Modelling such a transfer mechanism in our calculations (see

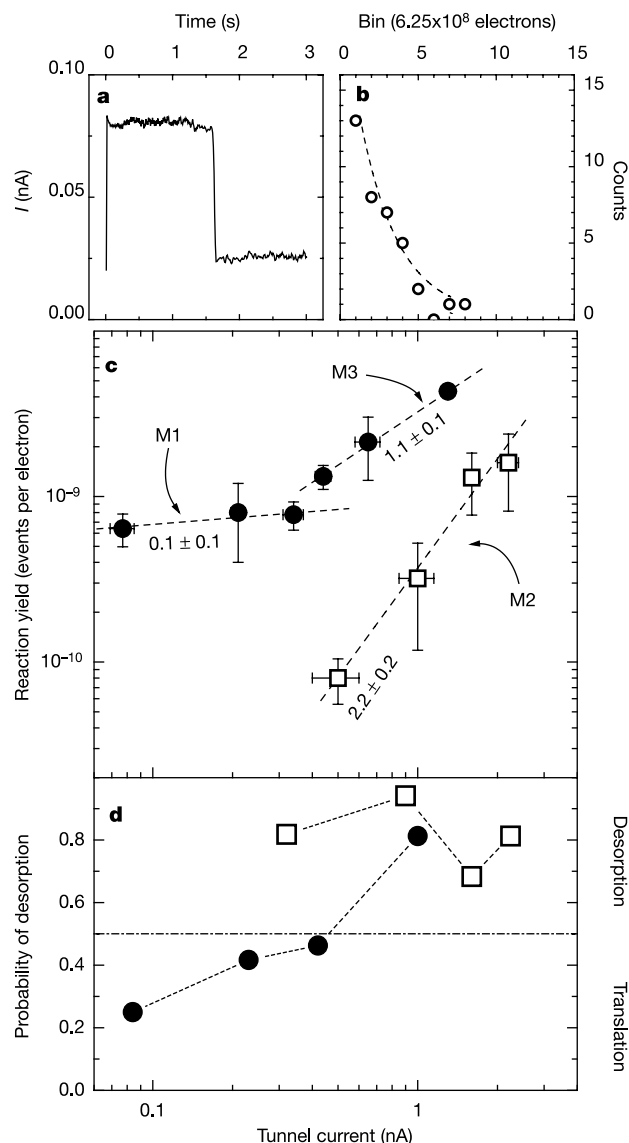


Figure 3 Dependence of reaction yield and product on the tunnelling current. Reactions are induced by current pulses with constant electron energy ($eV_s = 420\text{ meV}$ in **a**). **a**, The current drop shows when the reaction occurs. The area under the current pulse gives the number of electrons (n_e) used. **b**, The binned distribution of n_e shows an exponential decay ($V_s = 420\text{ mV}$, $I_t = 0.08\text{ nA}$). The reaction yield is the inverse of the exponential decay constant. **c**, Dependence of Y_R on the tunnelling current for reactions induced with tunnelling electrons of energy 420 meV (filled circles) and 320 meV (open squares). Dashed lines are least-square fits of points in regimes M1, M2 and M3, with slopes 0.1, 2.2 and 1.1 respectively. M1 and M2 indicate mechanisms identified after Fig. 2. **d**, Desorption probability P_{DES} as a function of tunnelling current for electron energies shown in **c**. The probability of molecular translation is $1 - P_{\text{DES}}$.

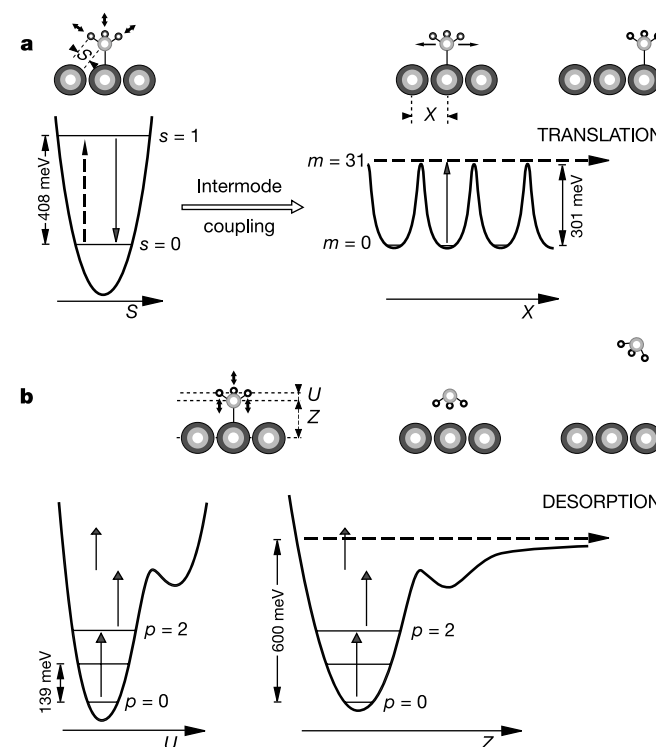


Figure 4 Scheme and one-dimensional potentials of reaction pathways M1 and M2. **a**, M1: 420-meV electrons excite the N-H stretch mode s (dashed arrow). Upon decay of the vibrational excitation, there is a finite probability of exciting hindered translational states m via anharmonic intermode coupling. Population of states $m > 30$ produces molecular motion above the 301-meV barrier for diffusion. **b**, M2: for 320-meV electrons the N-H stretch is a closed channel. Only the umbrella mode is operative. The two potential wells along the inversion coordinate (U) are no longer symmetric for the chemisorbed molecule. The inverted minimum is 360 meV above the umbrella state $p = 0$. This mode produces molecular motion along the adsorption coordinate Z , and thus steers desorption by inverting the ammonia molecule. The possible three excitations needed to overcome the 600-meV adsorption well are drawn.

Methods), we estimate that the probability of populating unbound translational states upon $\nu_s(\text{N-H})$ decay, and hence producing molecular translation, is 6×10^{-5} . Considering both factors (probability of $\nu_s(\text{N-H})$ excitation and probability of decay to translation), the small magnitude of the yield measured for pathway M1 is fully consistent with a mode-selective mechanism of inducing molecular translation, which prevails at lower tunnelling currents. When we raise the excitation rate in order to balance the stretch-to-translation decay rate with an additional excitation of the stretch mode, the energy accumulated in the molecule is large enough to activate an additional decay pathway producing molecular desorption, which is mediated by the umbrella mode (N.L. and J.I.P., manuscript in preparation).

Molecular desorption mediated by the $\delta_s(\text{N-H}_3)$ mode, M2, dominates when we reduce the electron energy to exclude $\nu(\text{N-H})$ excitations. This pathway is described by three excitations of the umbrella mode overtones in a ladder-climbing fashion (Fig. 4b). The ammonia gains sufficient vibrational energy to overcome the 600-meV adsorption well. The umbrella mode features a smaller coupling with translational states than that found for the $\nu_s(\text{N-H})$ mode. Moreover, our calculations show that $\delta_s(\text{N-H}_3)$ vibrational states above 360 meV may well populate a long-lived transition state connected with the complete inversion of the molecule, which leads to desorption after an additional excitation. In this process the umbrella mode itself becomes the vibrational state along the reaction coordinate (Fig. 4b), as was previously deduced from ultraviolet and infrared photodesorption experiments^{19,20}.

It would be interesting to extend this methodology to more complex processes, searching for strategies of controlling and enhancing reactivity at surfaces that may be applied at the macroscopic scale. The controlled environment furnished by the STM allows the detection of reaction mechanisms in the limit of very low yield and very low power irradiation. In such single-molecule studies, we expect that mode-selective strategies will become important in the discovery of reaction pathways that are inaccessible by classical 'thermal' chemistry. □

Methods

To estimate the potential barriers for translation and desorption, we performed total-energy calculations using plane waves and pseudopotentials in the generalized gradient approximation of density functional theory. Using the method of refs 23 and 24, we also estimated chemisorbed ammonia frequencies, the probability of excitation of each mode and their lifetime. The numerical results for the relevant modes of this work are: the symmetric and one antisymmetric stretch mode, $\nu_s(\text{N-H})$ and $\nu_a(\text{N-H})$, at 408 meV and 422 meV (294 meV and 311 meV for ND_3), both with a probability $P_\nu = 6 \times 10^{-4}$ of being excited per impinging electron and a lifetime of 4 ps and 8 ps, respectively; the umbrella mode, $\delta_s(\text{N-H}_3)$, at 139 meV (104 meV for ND_3), 40×10^{-4} and 25 ps; the scissors modes, $\delta_a(\text{N-H}_3)$, degenerate at 200 meV (145 meV for ND_3), 2×10^{-4} and 11 ps.

The decay probability, P_m , of a single $\nu_s(\text{N-H})$ excitation decaying into the unbound hindered translation levels plus electron-hole excitation is estimated as the fraction $w_{m>30}/w$. $w_{m>30}$ is the sum of damping rates into states m above the barrier for translation, and w is the damping rate of the $\nu_s(\text{N-H})$ mode, via excitation of electron-hole pairs and hindered translational states. An upper limit of the translation yield is then $Y_R = P_\nu \times P_m = 3 \times 10^{-8}$ molecules per electron.

Received 20 December 2002; accepted 8 April 2003; doi:10.1038/nature01649.

- Jortner, J., Levine, R. D. & Pullman, B. (eds) *Mode Selective Chemistry* (Kluwer Academic, Dordrecht, 1991).
- Dai, H. L. & Ho, W. (eds) *Laser Spectroscopy and Photochemistry on Metal Surfaces* (World Scientific, Singapore, 1995).
- Higgins, J., Conjusteau, A., Scoles, G. & Bernasek, S. L. State selective vibrational ($2\nu_3$) activation of the chemisorption of methane on Pt(111). *J. Chem. Phys.* **114**, 5277–5283 (2001).
- Potter, E. D., Herek, J. L., Pedersen, S., Liu, Q. & Zewail, A. H. Femtosecond laser control of a chemical reaction. *Nature* **355**, 66–68 (1992).
- Sinha, A., Hsiao, M. C. & Crim, F. F. Controlling bimolecular reactions: Mode and bond selected reaction of water with hydrogen atoms. *J. Chem. Phys.* **94**, 4928–4935 (1991).
- Bronikowski, M. J., Simpson, W. R., Girard, B. & Zare, R. N. Bond-specific chemistry: OD:OH product ratios for the reactions $\text{H} + \text{HOD}(100)$ and $\text{H} + \text{HOD}(001)$. *J. Chem. Phys.* **95**, 8647–8648 (1991).
- Hansma, P. K. (ed.) *Tunnelling Spectroscopy: Capabilities, Applications, and New Techniques* (Plenum, New York, 1982).
- Stipe, B. C., Rezaei, M. A. & Ho, W. Single-molecule vibrational spectroscopy and microscopy. *Science* **280**, 1732–1735 (1998).

- Ho, W. Inducing and viewing bond selected chemistry with tunneling electrons. *Acc. Chem. Res.* **31**, 567–573 (1998).
- Eigler, D. M., Lutz, C. P. & Rudge, W. E. An atomic switch realized with the scanning tunnelling microscope. *Nature* **352**, 600–603 (1991).
- Stipe, B. C., Rezaei, M. A. & Ho, W. Coupling of vibrational excitation to the rotational motion of a single adsorbed molecule. *Phys. Rev. Lett.* **81**, 1263–1266 (1998).
- Komeda, T., Kim, Y., Kawai, M., Persson, B. N. J. & Ueba, H. Lateral hopping of molecules induced by excitation of internal vibration mode. *Science* **295**, 2055–2058 (2002).
- Bartels, L. *et al.* Atomic scale chemistry: Desorption of ammonia from Cu(111) induced by tunneling electrons. *Chem. Phys. Lett.* **313**, 544–552 (1999).
- Avouris, Ph. Manipulation of matter at the atomic and molecular level. *Acc. Chem. Res.* **28**, 95–102 (1995).
- Stipe, B. C. *et al.* Single-molecule dissociation by tunneling electrons. *Phys. Rev. Lett.* **78**, 4410–4413 (1997).
- Hla, S. W., Bartels, L., Meyer, G. & Rieder, K. H. Inducing all steps of a chemical reaction with the scanning tunneling microscope tip: Towards single molecule engineering. *Phys. Rev. Lett.* **85**, 2777–2780 (2000).
- Kim, Y., Komeda, T. & Kawai, M. Single-molecule reaction and characterization by vibrational excitation. *Phys. Rev. Lett.* **89**, 126104 (2002).
- Rust, H.-P., Buiset, J., Schweizer, E. K. & Cramer, L. High precision mechanical approach mechanism for a low temperature scanning tunneling microscope. *Rev. Sci. Instrum.* **68**, 129–132 (1997).
- Hertel, T., Wolf, M. & Ertl, G. UV photostimulated desorption of ammonia from Cu(111). *J. Chem. Phys.* **102**, 3414–3430 (1995).
- Hussla, I. *et al.* Infrared-laser-induced photodesorption of NH_3 and ND_3 adsorbed on single-crystal Cu(100) and Ag film. *Phys. Rev. B* **32**, 3489–3501 (1985).
- Prybyla, J. A., Heinz, T. F., Misewich, J. A., Loy, M. M. T. & Glowacki, J. H. Desorption induced by femtosecond laser pulses. *Phys. Rev. Lett.* **64**, 1537–1540 (1990).
- Salam, G. P., Persson, M. & Palmer, R. E. Possibility of coherent multiple excitation in atom transfer with a scanning tunneling microscope. *Phys. Rev. B* **49**, 10655–10662 (1994).
- Lorente, N. & Persson, M. Theory of single molecule vibrational spectroscopy and microscopy. *Phys. Rev. Lett.* **85**, 2997–3000 (2000).
- Lorente, N. & Persson, M. Theoretical aspects of tunneling-current-induced bond excitation and breaking at surfaces. *Faraday Discuss.* **117**, 277–290 (2000).

Acknowledgements J.I.P. acknowledges research contracts 'Marie Curie' (EU) and 'Ramon y Cajal' (Ministerio de Ciencia y Tecnología). N.L. acknowledges support from ACI Jeunes Chercheurs, and the CNRS programme 'Nano-Objet Individuel'. All calculations were performed at the Centre d'Informatique Nationale de l'Enseignement Supérieur (CINES) and the Centre de Calcul Midi-Pyrénées (CALMIP).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.I.P. (pascual@icmab.es).

Impact of urbanization and land-use change on climate

Eugenia Kalnay & Ming Cai

University of Maryland, College Park, Maryland 20770-2425, USA

The most important anthropogenic influences on climate are the emission of greenhouse gases¹ and changes in land use, such as urbanization and agriculture². But it has been difficult to separate these two influences because both tend to increase the daily mean surface temperature^{3,4}. The impact of urbanization has been estimated by comparing observations in cities with those in surrounding rural areas, but the results differ significantly depending on whether population data⁵ or satellite measurements of night light^{6–8} are used to classify urban and rural areas^{7,8}. Here we use the difference between trends in observed surface temperatures in the continental United States and the corresponding trends in a reconstruction of surface temperatures determined from a reanalysis of global weather over the past 50 years, which is insensitive to surface observations, to estimate the impact of land-use changes on surface warming. Our results suggest that half of the observed decrease in diurnal temperature range is due to urban and other land-use changes. Moreover, our estimate of 0.27 °C mean surface warming per century due to

land-use changes is at least twice as high as previous estimates based on urbanization alone^{7,8}.

Two methods used in the US to classify meteorological stations into urban and rural to 'correct' the observed surface temperature trends for urbanization effects are based on population data⁵ and satellite measurements of night-light^{6–8}, respectively, and the corresponding estimates of the impact of urbanization differ in magnitude (0.06 and 0.15 °C per century)^{7,8}. The finding that atmospheric temperatures as measured by satellites and weather balloons have smaller warming trends than surface observations has been the subject of much discussion⁹ centred mostly on the quality of the data, but it could be partially explained by a predominance of land-use effects over greenhouse warming near the surface.

We estimated the impact of urbanization and other land uses on climate change by comparing trends observed by surface stations with surface temperatures derived from the NCEP-NCAR 50-year Reanalysis (NNR)¹⁰. In the NNR (a statistical combination of 6-hour forecasts and observations), surface observations of temperature, moisture and wind over land are not used¹¹. However, atmospheric vertical soundings of wind and temperature (rawinsondes and satellite soundings) strongly influence the NNR, and surface temperatures are estimated from the atmospheric values. As a result, the NNR should not be sensitive to urbanization or land-use effects, although it will show climate changes to the extent that they affect the observations above the surface.

As indicated by Fig. 1 and many other studies, the NNR captures well surface temperature variations caused by atmospheric storms, advection of warm/cold air, and variations in the frequency or track of major storms. In contrast to the actual surface observations, we find no statistically significant difference in the NNR estimation of urban and rural station trends (see Methods). These arguments suggest that we could attribute the differences between monthly or annually averaged surface-temperature trends derived from observations and from the NNR primarily to urbanization and other changes in land use.

We compare the daily maximum and minimum temperatures of 1,982 surface stations located below 500 m in the 48 contiguous United States, and the daily surface maximum and minimum temperatures on a 2.5° gaussian grid from the NNR interpolated to the station locations, both for the period 1950–1999. We compute temperature anomalies with respect to the 50-year mean annual cycle for each site and each data set. Trends are computed as changes in decadal averages in the anomalies to reduce random errors. The NNR (1948 to the present) has been constructed with a model and data assimilation system kept unchanged, but it is affected by changes in the observing systems, especially the introduction of the satellite observing system in 1979. Therefore, in the computation of trends we exclude changes from the decade of the 1970s to the 1980s.

Figure 1 compares time series of 50 years of monthly mean temperature anomalies for Baltimore, a large city in Maryland, including the average decadal difference between observations and the NNR. There is good agreement in the interannual variability, with a correlation of over 0.9, but also a growing trend in the difference between the surface observations and NNR, increasing to 1.4 °C during the 1990s, a difference we attribute to urbanization and other surface changes that do not affect the NNR. A similar analysis on all surface stations (Supplementary Fig. 1) indicates a 50-year correlation of about 0.9 everywhere except in mountainous regions, where it is between 0.4 and 0.7, which is why we only include stations located below 500 m. The correlation is also lower in the west coast, possibly owing to the proximity of mountains or to low data density in the Pacific Ocean. Decadal trends can be locally dominated by interannual and decadal variability of the temperature due to anomalies in the circulation rather than to land use change—effects that are excluded by taking the differences between surface and NNR temperatures.

The decadal trend averaged over the two separate 20-year periods (1980–1999, and 1960–1979) is computed for every station and averaged in boxes of 0.5° latitude by 0.5° longitude, with an overall average computed over all the boxes. Figures 2 and 3 show the average trends for the observations and from the NNR, and also the difference between these two trends, which is at least partially attributable to changes in use of the land surface.

The maximum temperature (Fig. 2) shows a warming trend in the observations in the eastern and western US and a cooling trend in the Midwest, with a slightly negative overall average of -0.017 °C per decade. The NNR is similar but smoother, with an average of $+0.008$ °C per decade. The difference between the observed and NNR trends is somewhat negative in most of the country east of the Rockies, but is strongly positive in California and to a lesser extent, in Oregon and Washington, with an average difference of -0.025 °C per decade.

The minimum temperature (Fig. 3) observations show a much

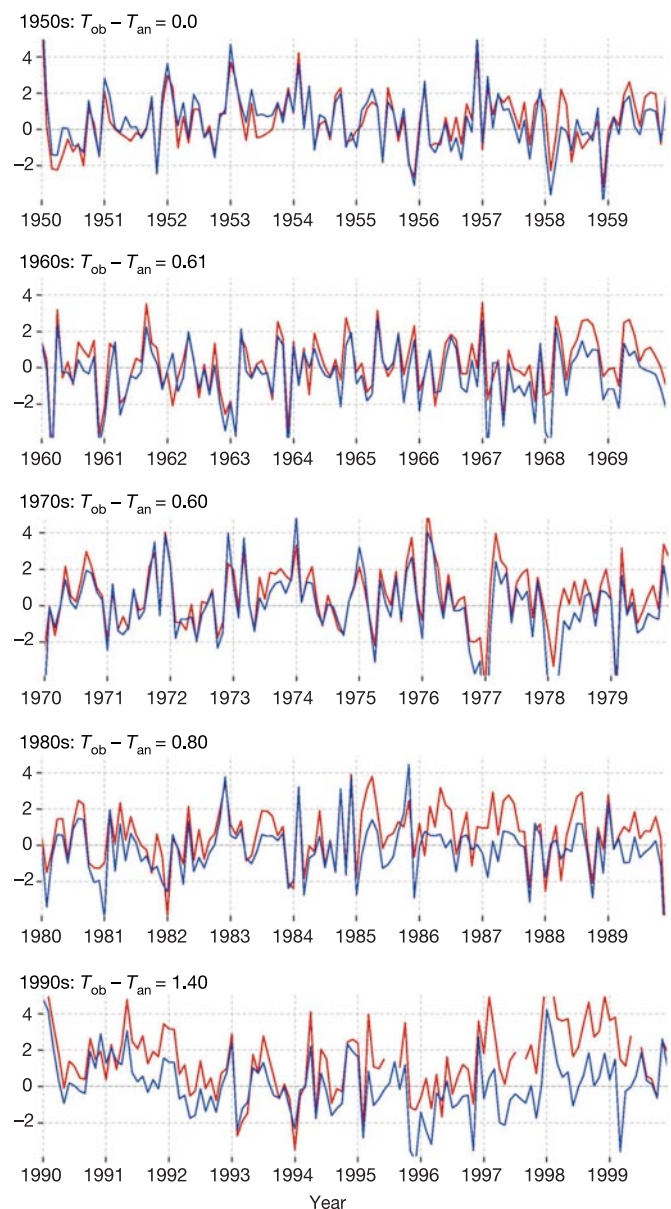


Figure 1 Comparison of monthly mean station and NNR surface temperature anomalies with respect to their annual cycles for the city of Baltimore, Maryland, USA. T_{ob} , observed monthly mean temperature in °C, shown in red. T_{an} , analysed monthly mean temperature in °C, shown in blue. Five decades (1950 to 1999) are shown for comparison.

stronger positive trend in most of the country, with an average of $+0.193^{\circ}\text{C}$ per decade. In the NNR, the minimum temperature increases everywhere except in the Midwest and California, with an average of $+0.113^{\circ}\text{C}$ per decade. The difference in minimum temperature trends between observed and NNR values is positive in most of the country, especially in California, with an average of 0.080°C per decade (40% of the observed trend).

Supplementary Fig. 2 shows the trend in the difference between maximum and minimum temperatures or diurnal temperature range (DTR). In the observations the DTR trend is strongly negative in most of the country with an average decrease of -0.210°C per decade. The NNR also shows a general decrease of DTR, with a national average of -0.105°C per decade so that about half of the decrease in DTR could be attributable to surface changes.

The daily mean temperature observation trends (Supplementary Fig. 3) obtained as the average of the maximum and minimum temperatures show an increase in most of the country, with an average trend of $+0.088^{\circ}\text{C}$ per decade. The NNR trends have an average of $+0.061^{\circ}\text{C}$ per decade. Of the two ‘urban correction’ estimates^{7,8}, our estimate of 0.27°C per century attributable to land use is closer to the estimate based on the night-light urban effect (see centre bottom of plate 3C in ref. 7) than to the estimate based on population density (see centre bottom of plate 3B in ref. 7). It should be noted that our observed daily mean temperature trends

(Supplementary Fig. 3a) are different from previous 50-year or 100-year trend estimates (see plates 3A and 7A in ref. 7) because in our computations we did not include (1) the decadal trends corresponding to the 1980s–1970s and especially the 1960s–1950s, and (2) urban and non-urban data adjustments. The non-urban adjustments tend to be strongly positive except over the Rockies (see plate 3B in ref. 7), so that if we had added them to the raw observations, our estimate of the land-use impact on the mean temperature trends would have been geographically similar but larger.

Although it is not possible definitively to attribute the differences between the observation and the NNR temperature trends solely to land use, including urbanization, agriculture and irrigation, our results are compatible with such an interpretation. The well-known ‘urban heat island’ effect actually takes place at night, when buildings and streets release the solar heating absorbed during the day. At the time of the maximum temperature the urban effect is one of slight cooling, owing to shading, aerosols, and to thermal inertia differences between city and country that are not currently well understood¹².

The effect of agricultural development, increasing evaporation during the day, would also tend to decrease the maximum temperature: irrigation would increase the heat capacity of the soil, thus increasing the minimum temperature. Therefore, both urbanization and agriculture effects could be consistent with the general increase in the minimum temperature and slight decrease in the maximum temperature, and contribute to the reduction in the diurnal temperature range shown in our estimates east of the Rockies (Figs 2c and 3c).

This implies that the comparison of urban and rural stations without including agricultural effects would underestimate the total

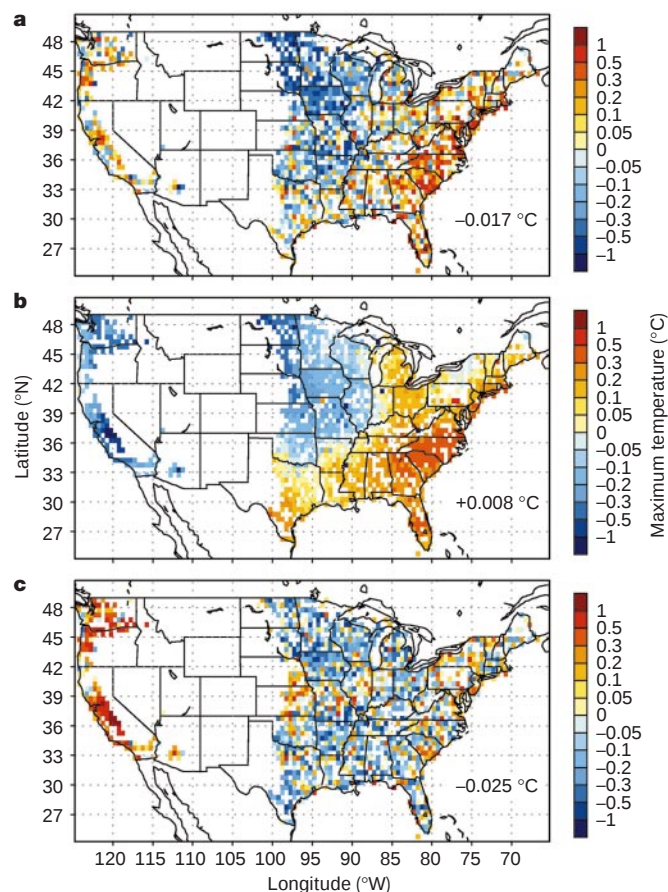


Figure 2 Decadal trend of the maximum temperature averaged for every US station below an elevation of 500 m. Each value (in $^{\circ}\text{C}$ per decade) was calculated from the average of the ‘1990s minus the 1980s’ and the ‘1970s minus the 1960s’ maximum temperatures. The station values are displayed as averages in boxes of 0.5° latitude by 0.5° longitude. Blank boxes indicate that none of the 1982 stations is within the boxes, and the national average is the average of these boxes. The average value of the trend is indicated in each panel. **a**, Station (observed) maximum temperature trends. **b**, NNR (analysed) maximum temperature trends. **c**, ‘Observed minus analysed’ maximum temperature trends.

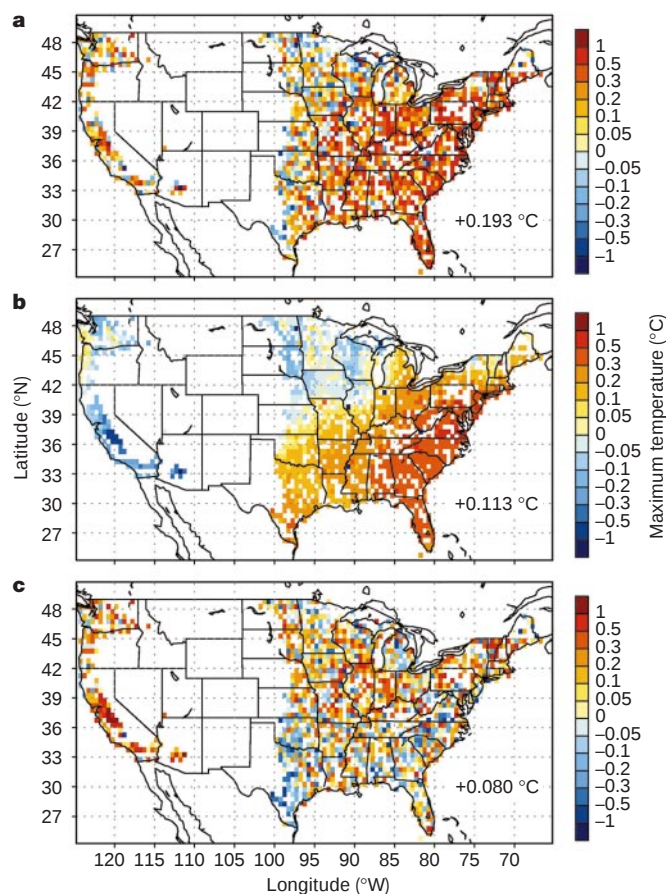


Figure 3 Decadal trend of the minimum temperature averaged for every US station below 500 m. Legend as for Fig. 2 but for the minimum temperatures.

impact of land-use changes. More studies are needed, including a comparison of geographical distribution of NNR trends with other upper-air observations, such as rawinsondes and satellites, a more precise definition of the urban and rural observing stations, and the impact of other human activities such as contrails and aerosols that can also reduce the diurnal temperature range¹³.

Our method can incorporate updated observations as they become available, can be applied to land stations throughout the world, to other variables such as humidity and winds, detect seasonal trends, and signal changes in station locations that are otherwise difficult to identify. □

Methods

Data

For the surface observations, we use the daily surface maximum and minimum uncorrected surface station temperatures from the National Climate Data Center (NCDC) 'Cooperative Summary of the Day' data set over the 48 contiguous states of the United States for 1950–1999. For the NNR, we use the global daily surface maximum and minimum temperatures gridded on 2.5° gaussian boxes, also for the period 1950–1999.

Analysis

We interpolate linearly the gridded NNR data to each observational site, and only consider the sites that have a total of at least 480 (whole) months of observations. In addition, because the NNR has surface heights different from those of the real locations, and extrapolations underground can introduce errors overwhelming the signal of the real trends (Supplementary Fig. 2), in the computation of the trends we only consider sites with elevations lower than 500 m. There are 1,982 US surface stations satisfying these two conditions. We obtain monthly means by averaging daily data; daily mean temperatures are obtained by averaging maximum and minimum temperatures, and daily temperature ranges by subtracting the minimum from the maximum temperature.

Because the NNR can have systematic differences with observations, especially near the surface, owing to deficiencies in the model forecast or the method of assimilation, we remove the 50-year monthly mean annual cycle for each site from both the observations and the NNR. We are thus comparing anomalies with respect to the 50-year mean annual cycle. In the results we present both comparisons of the 50-year time series and trends. The trends are computed as changes in decadal averages in order to reduce random errors. We only consider two decadal trends: the decade 1990–1999 minus 1980–1989, and 1970–1979 minus 1960–1969. We do not include in the trends the difference between the decades 1960–1969 and 1950–1959, because the observing system during the 1950s was considerably less reliable than in later decades, and it underwent significant scheduling changes during 1958 (ref. 11).

In addition, we have to address changes in the observing systems, especially the introduction of the satellite observing system (of which the most important is the TIROS-N Operational Vertical Sounder, TOVS) starting in 1979. These two major changes are the main reason why trends in the NNR need to be carefully estimated. We therefore do not include the changes 1980–1989 minus 1970–1979. The two decadal changes that we keep correspond to the 1990s minus 1980s (20 years with satellite data), and 1970s minus 1960s (20 years essentially without satellite data). Thus, when we average them we obtain decadal trends from two independent and largely homogeneous 20-year periods.

We compared the 1990s versus 1980s trend of 775 stations classified as urban versus 167 stations classified as rural. The mean surface temperature increased by 0.31 °C for the urban stations and 0.13 °C for the rural stations, with standard deviations of about 0.5 °C each. The difference between urban and rural warming, 0.18 °C, is significant at a 99% level of significance. The trends for the reanalysis station estimates are 0.26 °C for urban and 0.25 °C for rural, with standard deviations of about 0.22 °C, and the difference 0.01 °C between urban and rural is insignificant, showing that the NNR is insensitive to surface effects.

In the time series we compute the 1950–1959 average temperature difference between the NNR and the surface station at each station and subtract it from the NNR. This forces the two time series to have the same 10-year time average during the 1950s and is done for display but does not affect the computation of the trends or correlations.

Received 18 December 2002; accepted 23 April 2003; doi:10.1038/nature01675.

1. IPCC *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, UK, 2001).
2. Pielke, R. A. Sr *et al.* The influence of land-use change and landscape dynamics on the climate system: Relevance to climate-change policy beyond the radiative effects of greenhouse gases. *Phil. Trans. R. Soc. Lond. A* **360**, 1–15 (2002).
3. Gallo, K. P. & Owen, T. W. Satellite-based adjustments for urban heat island temperature bias. *J. Appl. Meteorol.* **38**, 806–813 (1999).
4. Owen, T. W., Gallo, K. P., Elvidge, C. D. & Baugh, K. E. Using DMSP-OLS light frequency data to categorize urban environments associated with US climate observing stations. *Int. J. Remote Sensing* **19**, 3451–3456 (1998).
5. Easterling, D. R. *et al.* Maximum and minimum temperature trends for the globe. *Science* **277**, 364–367 (1997).
6. Gallo, K. P., Easterling, D. R. & Peterson, T. C. The influence of land use/land cover on climatological values of the diurnal temperature range. *J. Clim.* **9**(11), 2941–2944 (1996).
7. Hansen, J. E. *et al.* A closer look at United States and global surface temperature change. *J. Geophys. Res.* **106**, 23947–23963 (2001).
8. Gallo, K. P., Owen, T. W., Easterling, D. R. & Jamason, P. F. Temperature trends of the US historical

- climatology network based on satellite-designated land use/land cover. *J. Clim.* **12**(5), 1344–1348 (1999).
9. NRC Board on Atmosphere Sciences and Climate Committee Panel, John M. Wallace, chair, pages 1–71. *Reconciling Observations of Global Temperature Change* (National Research Council, Washington DC, 2000).
10. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–431 (1996).
11. Kistler, R. *et al.* The NCEP/NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**, 247–267 (2000).
12. Runnalls, K. E. & Oke, T. R. Dynamics and controls of the near-surface heat island of Vancouver, BC. *Phys. Geogr.* **21**, 283–304 (2000).
13. Travis, D. J., Carleton, A. M. & Lauritsen, R. G. Contrails reduce daily temperature range. *Nature* **418**, 601 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This study was partially supported by a Risk Prediction Initiative grant. We are grateful to A. Senerini, who performed most of the computations, to J. E. Janowiak and W. Ebisuzaki who provided the data, and to R. Murnane, T. Oke, J. Hansen, E. Rasmussen, R. Pielke Sr, T. vonder Haar and Z. Li for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.K. (ekalnay@atmos.umd.edu).

Field sports and conservation in the United Kingdom

T. E. E. Oldfield, R. J. Smith, S. R. Harrop & N. Leader-Williams

Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, Kent CT2 7NS, UK

Many natural habitats exist on privately owned land outside protected areas¹, but few governments can afford to enforce or subsidize conservation of this biodiversity. Even in some developed countries, conservation subsidy schemes have only achieved limited success^{2–4}. Fortunately, some landowners may be willing to accept management costs in return for other benefits⁵, although this remains controversial when it involves the killing of charismatic species. For example, participants in British field sports, such as fox hunting and game-bird shooting, may voluntarily conserve important habitats that are required by quarry species^{6–8}. Here we report results from a multidisciplinary study that addressed this issue by focusing on three sites across central England. We found that landowners participating in field sports maintained the most established woodland and planted more new woodland and hedgerows than those who did not, despite the equal availability of subsidies. Therefore, voluntary habitat management appears to be important for biodiversity conservation in Britain. Current debates on the future of field sports in Britain, and similar activities globally, may benefit from considering their utility as incentives to conserve additional habitat on private land.

Private landowners play an increasingly important role in biodiversity conservation¹. This is especially important where habitats form isolated remnants in an agricultural matrix, and it is politically difficult to establish large protected areas⁹. This is typified by the situation in Britain, where farmland covers 76% of the country and increases in agricultural efficiency have caused great declines in biodiversity^{7,10,11}. The British government has responded by introducing legislation to protect important habitats and species on public and private land^{12–14}, as well as establishing subsidy schemes^{11,15}. However, conservation legislation remains unpopular

with certain landowners, making enforcement by the statutory agencies difficult¹⁶. In contrast, subsidised agri-environment schemes do not involve coercion, but receive little funding¹⁷ and can be poorly targeted^{4,18}. Nevertheless, many British landowners are interested in conservation and may be willing to accept the costs of maintaining biodiversity¹⁹. However, research on this topic requires consideration of the role of field sports, which are controversial because of associated animal welfare issues.

Woodland and hedgerows are important habitats and create linkages across agricultural landscapes²⁰, while also providing important cover for British quarry species⁸. Both habitat types have declined considerably in the past 50 years^{21,22}, but those elements with high scenic or conservation value now have legal protection through a range of prescriptive legislation^{12,13}. In addition, particular subsidies now encourage maintenance and planting of woodland and hedgerows¹⁵, although funding availability is limited¹⁷. Moreover, uptake of such schemes depends on landowners' conservation values¹⁹ and field sports may play a role in maintaining these habitats⁸. But previous studies have either used self-selecting questionnaires^{6,7} or focused on the effects of game management practices on specific taxa^{23,24}. We therefore sought to determine whether those who hunt foxes and/or maintain a game-bird shoot on their land voluntarily increase the biodiversity value of their land. We measured the extent of woodland and hedgerows in three study sites in central England, and investigated whether participation in these field sports, as well as farm size, farm type, dependence on income from farming, and membership of a biodiversity advisory group, influenced a landowner's likelihood of conserving habitat (see Methods).

Analysis of aerial photographs showed that landowners who hunt and those who maintain game-bird shooting support more woodland cover than those not involved in field sports (hunt: F -test statistical datum $F = 4.004$, $P = 0.05$; shoot: $F = 12.439$, $P = 0.001$), with no effect of study site, farm size or type, income dependence, or advisory-group membership. Landowners who both hunted and maintained game-bird shoots conserved the most woodland cover, around 7% of their farm area (Fig. 1). There was no difference in the proportion of field boundaries consisting of woodland or hedgerow between landowners practising and not involved in field sports (hunt: $F = 0.305$, $P = 0.583$; shoot: $F = 0.956$, $P = 0.332$).

Interviews revealed that landowners who participated in hunting and shooting were more likely to have planted woodland (Area under the Receiver Operating Characteristic curve (AUC) = 0.781; hunt: Wald = 6.050, $P = 0.014$; shoot: Wald = 8.463, $P = 0.004$; Fig. 2), with no effect of study site, farm size or type, or advisory-group membership. These results support an earlier questionnaire

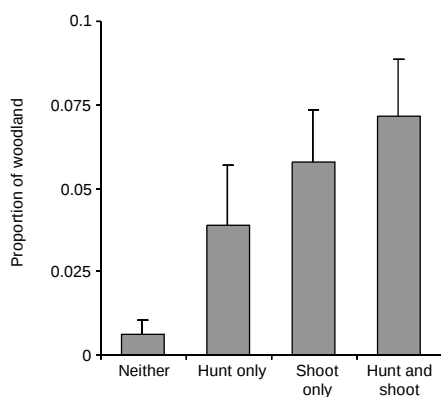


Figure 1 Influence of participation in field sports on the proportion of each farm covered with woodland. Data are mean $\pm$ s.e.

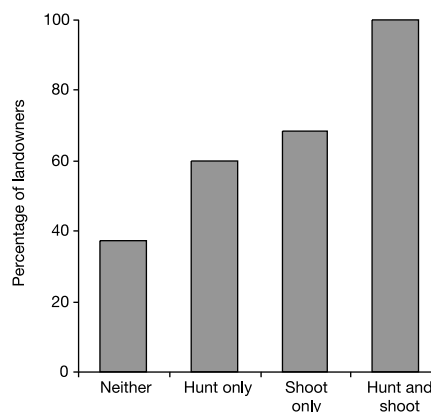


Figure 2 Influence of participation in field sports on the proportion of landowners planting new woodland.

study⁷, suggesting that involvement in field sports is an important incentive for farmers to create additional woodland. The likelihood of planting will also depend on the landowner's income, as new planting incurs direct and opportunity costs, even with grant support. Indeed, landowners who hunt and maintain game-bird shoots were less dependent on farming income (hunt: $\chi^2 = 4.197$, $P = 0.04$; shoot: $\chi^2 = 7.565$, $P = 0.006$), but wealth alone does not appear sufficient to encourage landowners to plant.

New hedgerow planting was predicted by hunting and advisory-group membership (AUC = 0.752; hunt: Wald = 6.166, $P = 0.013$; advisory group: Wald = 10.657, $P = 0.001$; Fig. 3), with no effect of study site, farm size or type, income dependence, or maintaining a game-bird shoot. Furthermore, hedgerows were generally richer in woody plant species when landowners belonged to an advisory group, although this depended on study site (site $\times$ advisory group: $F = 3.637$, $P = 0.007$; Fig. 4), with no effect of farm size or type, income dependence, or participation in field sports. Previous studies have shown that species richness in hedgerows is affected by both hedgerow age and current management²⁵. Hence, our findings further emphasize¹⁹ the value of landowners belonging to an advisory group for maintaining species diversity in hedgerows. Moreover, our findings suggest that participating in fox hunting can indirectly support hedgerow conservation through new planting, and supplement the direct support shown by landowners belonging to advisory groups.

Our results suggest that governments in developed countries, such as Britain, could benefit from adopting the sustainable-use and incentive-based conservation policies that they encourage

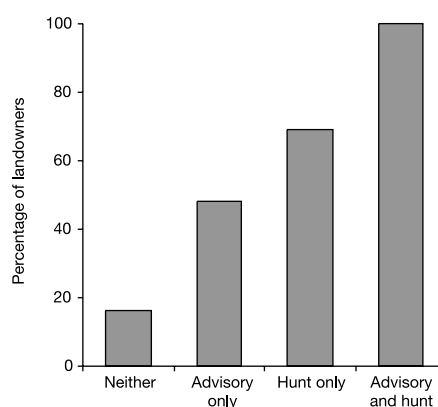


Figure 3 Influence of participation in hunting and membership of an advisory group on the proportion of landowners planting new hedgerows.

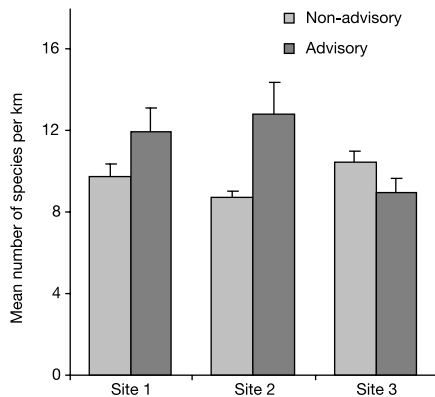


Figure 4 Influence of study site and membership of an advisory group on number of woody plant species in hedgerows. Data are mean $\pm$ s.e.

abroad^{26,27}. We have shown that landowners participating in British field sports are more likely to maintain established woodland habitat on their farms. More importantly, they are also more likely to undertake new plantings, even though all the farmers have equal opportunities to apply for subsidies that support these activities. Game-bird shooting produced a greater effect on established and planted woodland (Figs 1 and 2), which is unsurprising given the considerable financial return it can generate for landowners. Nevertheless, landowners who hunt with hounds are more likely to conserve woodland habitat (Fig. 1) and plant more woodland and hedgerows (Figs 2 and 3), suggesting that the perceived recreation and social benefits of this controversial activity can produce conservation benefits. However, current debate over the future of fox hunting with hounds predominantly focuses on welfare issues and its uncertain role in population control^{28,29}, leading to proposed legislation that seeks to balance cruelty against utility of control³⁰. Our results suggest an equally valid test of utility could focus on the role of landowners in voluntary habitat conservation. Equally, should hunting, or indeed game-bird shooting, be banned on welfare grounds without concessions for such utility, then additional public funds may be needed to increase subsidies for habitat conservation, together with the strengthened capacity to enforce legislation. □

Methods

Sampling protocol

Three study sites that each fell within one hunt country (an area hunted by an individual hunt) were chosen to represent arable, mixed and pastoral farming areas in central England, an area that has little coverage of formally protected areas. Lists of farmers were obtained at each site from the local foxhunt. Samples of hunting and non-hunting farmers were chosen using a random number generator. All selected farmers agreed to participate in the study, and this produced a total sample size across the three study sites of 65 landowners who owned more than half their farms and had farmed there for 10 years or more. Questionnaire-based interviews conducted with each landowner sought details of the following: farm boundaries; farm type (whether arable, livestock or mixed); farm income dependence (whether or not dependent solely on income from farming); non-productive land management (whether or not new woodland and hedgerows had been planted in the previous 10 years); advisory-group membership (whether or not the landowner belonged to an advisory group such as the Farming and Welfare Advisory Group); and maintenance of game shooting (whether or not farms maintained commercial or non-commercial game shooting on their land). Each selected farm was digitized from 1999/2000 aerial photos, and ArcView v3.2 GIS software (ESRI, Redlands, California) was used to determine its size, area of woodland and length of hedgerows. Hedgerow surveys—undertaken along 1.6 km of hedgerow from eight hedges randomly selected from digitized maps on each sampled farm—sought to determine the number of woody plant species in each hedge.

Statistical analysis

We sought to identify the factors that determine whether landowners conserved woodland and hedgerow habitat, irrespective of study site. Habitat conservation was assessed through five measures on a total of 65 farms: the proportion of each farm covered in established woodland; the proportion of landowners planting new woodland; the

proportion of farm boundary consisting of hedgerow and woodland on each farm; the proportion of landowners planting new hedgerows; and, the number of woody plant species per kilometre of hedgerow on each farm. Each of these dependent variables was compared against the following explanatory variables: study site; farm size; farm type; farm income dependence; advisory group member; participation in fox hunting; and maintenance of game shooting on the farm. All of these variables, apart from farm size, were categorical. General linear modelling was used to find the factors determining the following: the proportion of woodland; the proportion of farm boundary consisting of hedgerow and woodland; and the number of woody plant species per kilometre of hedgerow. In order to meet the assumptions of the general linear model, a square-root transformation was applied to the proportion of woodland. Stepwise logistic regression modelling was used to find the factors that determined the probability of landowners having planted woodland and hedgerows. The possible influence of spatial autocorrelation was investigated for each dependent variable within each study site by calculating the Moran's *I* statistic using the CrimeStatII software package (v2.0, Ned Levine & Associates, Houston, Texas) and no significant effect was evident.

Received 11 March; accepted 27 March 2003; doi:10.1038/nature01678.

- Langholz, J. A. & Lassoie, J. P. Perils and promise of privately owned protected areas. *BioScience* **51**, 1079–1085 (2001).
- Ovenden, G. N., Swash, A. R. H. & Smallshire, D. Agri-environment schemes and their contribution to the conservation of biodiversity in England. *J. Appl. Ecol.* **35**, 955–960 (1998).
- Krebs, J. R., Wilson, J. R., Bradbury, R. B. & Siriwardena, G. M. The second silent spring? *Nature* **400**, 611–612 (1999).
- Kleijn, D., Berendse, F., Smit, R. & Gilissen, N. Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. *Nature* **413**, 723–725 (2001).
- Goodman, P. S., James, B. & Carlisle, L. in *Mainstreaming Biodiversity in Development. Case Studies from South Africa* (eds Pierce, S. M., Cowling, R. M., Sandwith, T. & Mackinnon, K.) 21–32 (World Bank Environment Department, Washington DC, 2002).
- Macdonald, D. W. *Running with the Fox* (Oxford Univ. Press, Oxford, 1987).
- Macdonald, D. W. & Johnson, P. J. Farmers and the custody of the countryside: Trends in loss and conservation of non-productive habitats 1981–1998. *Biol. Conserv.* **94**, 221–234 (2000).
- Tapper, S. (ed.) *A Question of Balance; Game Animals and their Role in the British Countryside* (Game Conservancy Trust, Fordingbridge, UK, 1999).
- Leader-Williams, N., Harrison, J. & Green, M. J. B. Designing protected areas to conserve natural resources. *Sci. Prog. Oxford* **74**, 189–204 (1990).
- Robinson, R. A. & Sutherland, W. J. Post-war changes in arable farming and biodiversity in Great Britain. *J. Appl. Ecol.* **39**, 157–176 (2002).
- UK Government *Achieving a Better Quality of Life—Government Annual Report 2001* (DEFRA, London, 2001).
- The Town and Country Planning (Trees) Regulations 1999* Statutory Instrument 1999 No. 1892 (Controller of HMSO, London).
- The Hedgerows Regulations 1997* Statutory Instrument 1997 No. 1162 (Controller of HMSO, London).
- O'Connell, M. O. & Yallop, M. Research needs in relation to the conservation of biodiversity in the UK. *Biol. Conserv.* **103**, 115–123 (2002).
- DEFRA (Department for Environment, Food and Rural Affairs) (<http://www.defra.gov.uk/erdp/schemes/default.htm>) (posted 10 December 2002, modified 16th April 2003).
- English Nature *Annual Report April 1 2001–31 March 2002* (English Nature, Peterborough, UK, 2002).
- Lovelace, D., May, R. & Perkins, R. *Money Makes the Countryside Go Round* (WWF, Godalming, UK, 2000).
- Sutherland, W. J. Restoring a sustainable countryside. *Trends Ecol. Evol.* **17**, 148–150 (2002).
- Beedell, J. D. C. & Rehman, T. Explaining farmers' conservation behaviour: Why do farmers behave the way they do? *J. Environ. Mgmt* **57**, 165–176 (1999).
- Bennett, A. F. *Linkages in the Landscape: The Role of Corridors in Wildlife Conservation* (IUCN, Gland, Switzerland, 1999).
- Thomas, R. C., Kirby, K. J. & Reid, C. M. The conservation of a fragmented ecosystem within a cultural landscape—the case of ancient woodland in England. *Biol. Conserv.* **82**, 243–252 (1997).
- Barr, C. J. *et al. Countryside Survey 1990 Main Report* (Department of Environment, London, 1993).
- Robertson, P. A., Woodburn, M. I. A. & Hill, D. A. The effects of woodland management on the abundance of butterflies in Dorset, England. *Biol. Conserv.* **45**, 159–167 (1988).
- Stoate, C. Multifunctional use of a natural resource on farmland: Wild pheasant (*Phasianus colchicus*) management and the conservation of farmland passerines. *Biodivers. Conserv.* **11**, 561–573 (2002).
- Garbutt, R. A. & Sparks, T. H. Changes in the botanical diversity of a species rich ancient hedgerow between two surveys (1971–1998). *Biol. Conserv.* **106**, 273–278 (2002).
- Glowka, L., Burhenne-Guilman, F. & Synge, H. A *Guide to the Convention on Biological Diversity* (IUCN, Gland, Switzerland, 1994).
- Leader-Williams, N., Oldfield, T. E. E., Smith, R. J. & Walpole, M. J. Science, conservation and fox-hunting. *Nature* **419**, 878 (2002).
- Burns, L., Edwards, V., Marsh, J., Soulsby, L. & Winter, M. *Report of the Committee of Inquiry into Hunting with Dogs in England and Wales* (Stationery Office, London, 2000).
- Baker, P. J., Harris, S. & Webbon, C. C. Effect of British hunting ban on fox numbers. *Nature* **419**, 34 (2002).
- The Hunting Bill 2002 (<http://www.publications.parliament.uk/pa/cm200203/cmbills/010/03010.i-v.html>) (13 December 2002).

Acknowledgements We thank farmers in the Suffolk, Warwickshire and Berkeley hunt countries for their co-operation, and M. J. Walpole and M. S. Ridout for advice. This work was supported by CHK Charities.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.L.-W. (N.Leader-Williams@uk.ac.uk).

Action video game modifies visual selective attention

C. Shawn Green & Daphne Bavelier

Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627, USA

As video-game playing has become a ubiquitous activity in today's society, it is worth considering its potential consequences on perceptual and motor skills. It is well known that exposing an organism to an altered visual environment often results in modification of the visual system of the organism. The field of perceptual learning provides many examples of training-induced increases in performance. But perceptual learning, when it occurs, tends to be specific to the trained task; that is, generalization to new tasks is rarely found^{1–10}. Here we show, by contrast, that action-video-game playing is capable of altering a range of visual skills. Four experiments establish changes in different aspects of visual attention in habitual video-game players as compared with non-video-game players. In a fifth experiment, non-players trained on an action video game show marked improvement from their pre-training abilities, thereby establishing the role of playing in this effect.

We first used the flanker compatibility effect, a standard experimental paradigm in attentional studies, to determine whether video-game playing produces an overall increase in attentional capacity¹¹. This task measures the effect of a to-be-ignored distractor on a target task. The size of this distractor effect has been proposed to provide an index of 'left-over' attentional resources. In particular, it has been reported¹² that the distractor effect is large when the target task is easy, but small when the target task is made difficult by increasing the number of filler shapes in the ring (Fig. 1a). The currently accepted basis for this finding is that when the target task is easy, spare attentional resources 'spill over' to the distractor, processing it to some extent and thereby influencing target processing. As the target task becomes more difficult, fewer attentional resources remain to process the extraneous distractor. Irrelevant processing is only truly prevented once the target task becomes so difficult that all available resources are devoted to the target task, leaving none to the distractor.

This task therefore affords an opportunity to test the hypothesis that video-game playing increases the capacity of the visual attentional system. If video-game players (VGPs) do indeed have a greater attentional capacity, they should exhaust their visual attention resources more slowly than non-video-game players (NVGPs) as the target task becomes more difficult. As predicted by the hypothesis of a greater capacity in their population, VGPs showed an overall distractor effect that remained even when the target task was difficult (Fig. 1b). Thus, at difficulty levels where NVGPs have long depleted their attentional resources, VGPs possess sufficient resources to perform the target task, with resources still remaining to spill over to the distractors. These results indicate that VGPs possess enhanced attentional capacity.

We confirmed this increase in capacity more directly through the use of an enumeration task. When asked to report how many squares are presented in a briefly flashed display, normal subject performance seems to be best captured by two distinct processes: an 'automatic' operation for small numbers of targets in which performance is rapid, accurate and independent of number (subitizing); and a slower, serial counting process for larger numbers of items¹³. The number of items that can be subitized gives an estimate of the number of items that can be attended at once^{14–16}. VGPs could subitize more items than NVGPs (4.9 versus 3.3 items; population effect, $P < 0.01$; Fig. 2). Thus,

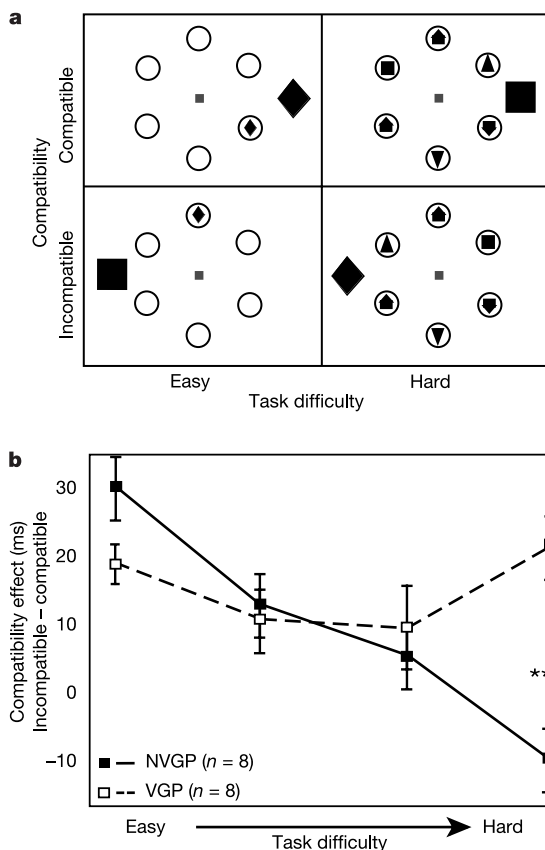


Figure 1 Measure of attentional resources. **a**, Flanker compatibility task. Participants were asked to decide whether a square or a diamond appeared within one of the six rings (target task), while ignoring a distractor shape presented outside the rings. The distractor shape could be either compatible (the same as the target shape) or incompatible (the alternative target shape). Of interest was the difference in target processing speed between compatible and incompatible trials. This difference, also called the 'compatibility effect', measures the attentional resources available to the participant. By inserting no, one, three or five extra shapes into the rings, the difficulty of the target task was manipulated to test the possibility that VGPs show compatibility effects at task difficulties for which attentional resources are usually exhausted in NVGPs. **b**, Flanker compatibility effect. As previously reported¹², the compatibility effect (and spare attentional resources) decreased with task difficulty in NVGPs (task difficulty $\times$ compatibility, $P < 0.001$). In VGPs, by contrast, the compatibility effect remained as the target task was made more difficult (population $\times$ task difficulty $\times$ compatibility, $P < 0.003$). As a result, VGPs showed a significantly greater compatibility effect than NVGPs when the target task was difficult, indicating greater attentional resources in VGPs (** $P < 0.01$). Error bars denote s.e.m.

video-game playing enhances the number of visual items that can be unerringly apprehended.

These two experiments indicated that video-game playing enhances attentional capacity in an area well within the 'training' (video-game playing) zone (0–5° from fixation); however, it was unclear whether video-game playing also facilitates processing outside the training range. To test this possibility, we compared the distribution of visual attention in VGPs and NVGPs at three locations, one well within the training range (10°), one at its boundary (20°) and one well outside the training range (30°; all VGPs reported playing in a field that extends no more than 18° from fixation). We adapted the 'useful field of view' task (refs 17–19 and Fig. 3a), which measures participants' ability to locate a target amongst distractors. This measure does not correlate well with standard tests of visual acuity but rather provides a measure of attentional resources and their spatial distribution¹⁷. VGPs far outperformed NVGPs at all eccentricities (Fig. 3b), distractor levels and distractor–eccentricity pairs. These results indicate an

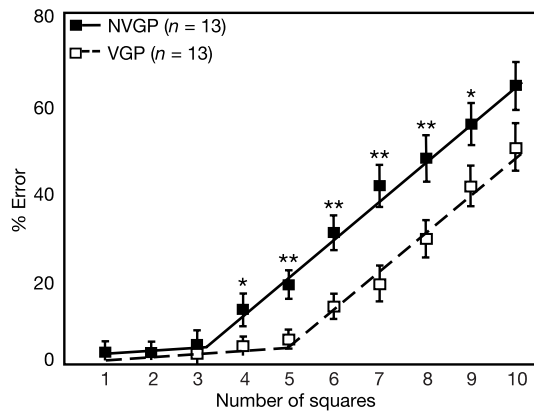


Figure 2 Enumeration performance. When asked to report the number of squares briefly flashed, VGPs were able to apprehend more items at once than were NVGPs (4.9 versus 3.3). Overall, VGPs were significantly more accurate than NVGPs (78% versus 65%, $P < 0.003$) and, as expected, this difference only emerged for numbers above the subitizing range of NVGPs. Error bars denote s.e.m. (* $P < 0.05$, ** $P < 0.01$).

enhanced allocation of spatial attention over the visual field, even at untrained locations, in VGPs.

The three experiments described so far indicated that video-game playing enhances the capacity of visual attention and its spatial distribution. We next examined the temporal characteristics of visual attention and asked whether the pressure to act rapidly on several visual items, which is inherent to most action games, alters

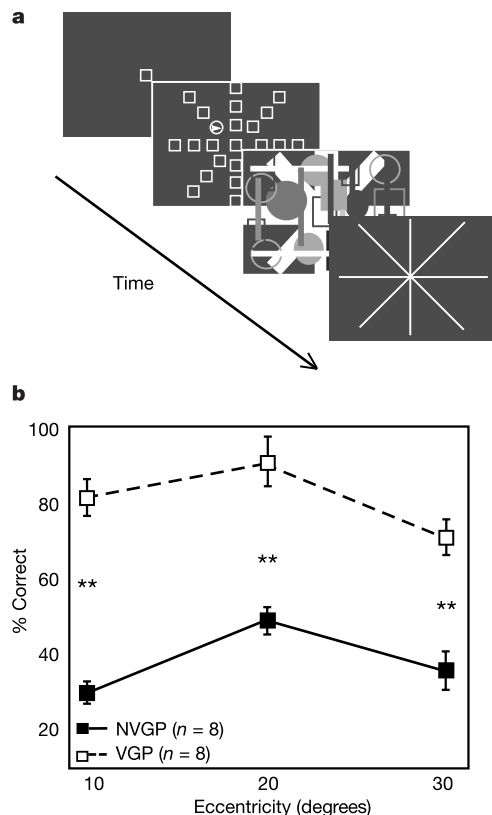


Figure 3 Measure of attention over space. **a**, Sequence of displays in the useful field of view task. Participants indicated the spoke on which the small target (triangle within circle) appeared. By presenting the target at different eccentricities, this task affords an opportunity to test the spatial distribution of visual attention. **b**, Localization accuracy. VGPs showed large enhancements in localization ability at all eccentricities. The superiority of VGPs at 30° indicates that the enhancement of spatial attention observed in this population is not limited to trained locations. Error bars denote s.e.m. (** $P < 0.01$).

the ability to process items over time, particularly the ability to avoid 'bottlenecks' of attention that often occur in temporal processing. The experimental paradigm that we used to test the temporal aspects of visual attention is the attentional blink task (refs 20–22 and Fig. 4a). Attentional blink refers to the phenomena wherein subjects have difficulties reporting a second target when it appears a few hundreds of milliseconds after the onset of a first target. In our case, we used a variant of the attentional blink, in which participants were asked to identify a first target and then detect a second target. This task includes two distinct attentional bottlenecks. First is the attentional blink *per se*, which is the difficulty of processing a second target that comes 200–500 ms after the onset of the first one. Second is the cost of switching tasks between the first and second target (from identification to detection); unlike the attentional blink *per se*, this effect is most pronounced when the two targets are temporally adjacent and then decreases slowly as the time between the two targets increases. This attentional bottleneck is not specific to vision but rather appears amodal^{21,23,24}. Thus, by using an identification/detection attentional blink task, we could test whether the enhanced capacities after video-game training not only applied to a purely visual bottleneck but also generalized to an amodal one.

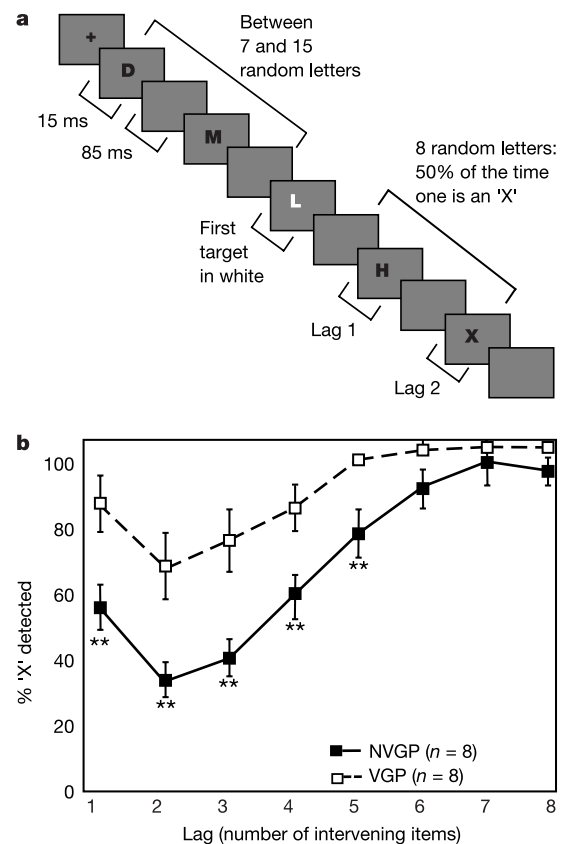


Figure 4 Measure of attention over time. **a**, Attentional blink task. Black letters were rapidly presented at fixation in a standard rapid, serial, visual presentation manner. At a random time in the stream, a white letter was presented (first target). After this first target, an 'X' (second target) was presented among the following letters 50% of the time. After the trial, the subject gave the identity of the first target and then indicated whether the 'X' was presented. Of interest is the performance of subjects on the 'X' detection task, given that they have correctly identified the white letter. **b**, Attentional blink performance. At early lags, VGPs performed better (less blink) than NVGPs; as lag increased, the effect of the attentional bottlenecks decreased and, as expected, the two populations became comparable (lag by population $P < 0.02$). Error bars denote s.e.m. Points without error bars indicate that the s.e.m. was smaller than the size of the square (** $P < 0.01$).

VGP's outperformed NVGP's on second-target correct detections from lag 1 to lag 5, indicating less attentional blink (Fig. 4b). The finding of a population difference as early as lag 1 indicates that video-game training enhances task-switching abilities as well as decreasing the attentional blink. Thus, both the visual and amodal bottlenecks identified during temporal processing of visual information are reduced in VGP's. Clearly, these individuals have an increased ability to process information over time; however, whether this is due to faster target processing, such as faster selection and stabilization of information in memory, or to an increased ability to maintain several attentional windows in parallel, cannot be determined from our current data.

We also carried out a training experiment to alleviate concerns over the source of the VGP and NVGP population differences. By selecting VGPs, we may have selected individuals with inherently better attentional skills than NVGP's. The consequence of this potential split would be that VGPs, with their greater natural ability, did well at video games and so played often, whereas NVGP's, whose inherent abilities limited their success, avoided playing games as a result. Another possible confound was that, because of relatively superior visuo-motor coordination, VGPs were better able to perform the motor aspect of the tasks and thus had more spare resources to devote to the visually demanding part of the tasks.

To rule out these potential confounds, a group of NVGP's underwent action-video-game training, in which they were asked to play *Medal of Honor* (a game similar to those reported being played by

the VGP population) for 1 h per day for 10 consecutive days. A control group was trained, over the same time span, on the game *Tetris*. This game contains a challenging visuo-motor component but, whereas action games require that attention is distributed and/or switched around the field, *Tetris* demands focus on one object at a time. *Tetris*, therefore, would not be expected to change the aspects of visual attention described above and thus affords an excellent control for monitoring improvement due to enhanced visuo-manual expertise and test-retest improvements.

Before and after training we tested each group on the enumeration, useful-field-of-view and attentional-blink experiments. The training was successful as all participants improved their scores on the video game on which they were trained. Notably, action-game training led to greater performance improvement than did the control game on all three experimental tasks. Enumeration increased by 1.7 items in individuals trained on action games, whereas the control group showed no improvement (population by improvement, $P < 0.05$). Training on action video games enhanced the useful field of view and lead to faster recovery from the attentional blink (Fig. 5). In addition, we found that the improvement in these attentional measures was marginally correlated with improvement in video-game playing (adjusted $r^2 = 0.43$, $P = 0.13$; see Supplementary Information). Thus, 10 days of training on an action game is sufficient to increase the capacity of visual attention, its spatial distribution and its temporal resolution.

By forcing players to simultaneously juggle a number of varied tasks (detect new enemies, track existing enemies and avoid getting hurt, among others), action-video-game playing pushes the limits of three rather different aspects of visual attention. It leads to detectable effects on new tasks and at untrained locations after only 10 days of training. Therefore, although video-game playing may seem to be rather mindless, it is capable of radically altering visual attentional processing. There are several ways by which video-game training could lead to such enhancements. Changes in known attentional bottlenecks is certainly a possibility; however, speeded perceptual processes and/or better management of several tasks at the central executive level are also likely to contribute. It will be for future studies of the effect of video-game practice to determine the relative contribution of these different factors to skill learning. □

Methods

Participants

Subjects were aged between 18 and 23 years. The VGPs had played action video games on at least 4 days per week for a minimum of 1 h per day for the previous 6 months. The games included *Grand Theft Auto3*, *Half-Life*, *Counter-Strike*, *Crazy Taxi*, *Team Fortress Classic*, *007*, *Spider-Man*, *Halo*, *Marvel vs Capcom*, *Roguespear* and *Super Mario Cart*. The NVGP's had little, and preferably no, video-game usage in the past 6 months. Experiments 1–4 included only males; in experiment 5, both male and female NVGP's underwent training.

Apparatus, stimuli and procedure

For experiment 1—flanker compatibility (Fig. 1)—the apparatus, stimulus, procedure and data analysis have been described²⁵.

For experiment 2—enumeration (Fig. 2)—the stimuli consisted of a random number (between 1 and 12) of white squares (each subtending $0.5^\circ \times 0.5^\circ$) presented for 50 ms within an invisible $5^\circ \times 5^\circ$ square centred on the screen. Subjects were asked to type the number of squares that they believed were presented; no feedback was provided. We scored the percentage of trials in which participants correctly reported the number of squares presented. Only data from trials with 1–10 squares are presented, as trials with 11 and 12 squares were excessively difficult for all populations. To estimate the point at which subjects switched from subitizing to counting, we modelled the data of each subject with a bilinear curve. The first line was set to have an intercept of (0,0) and was expected to have a rather shallow slope (subitizing range); the second line was expected to have a steeper slope (counting). The best bilinear fit was found by minimizing the least-square error. The point at which the two lines met was then computed and considered to be the point where subitizing switches to counting.

For experiment 3—useful field of view (Fig. 3)—the stimuli and procedure were similar to those used previously¹⁷, except that the stimulus size and presentation time were both decreased to account for the increased ability of comparatively younger subjects (target stimulus was a filled triangle within a $3^\circ \times 3^\circ$ circle outline; distracting squares subtended $4^\circ \times 4^\circ$, no central task). Separate timings were used to ensure equal difficulty at each

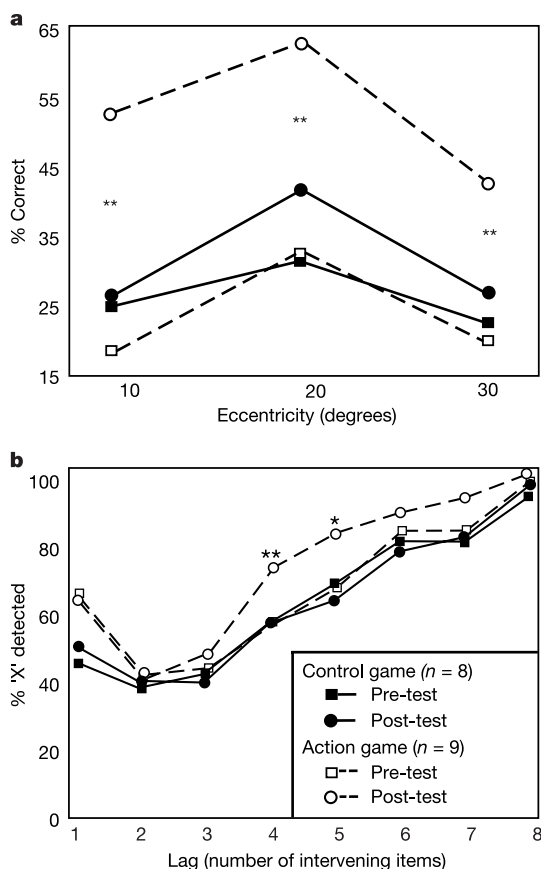


Figure 5 Performance before and after training. **a**, Training: useful field of view. At each eccentricity, the group trained on an action video game improved significantly more from their pre-test scores than did the control group trained on a non-action video game ($**P < 0.01$). **b**, Training: attentional blink. The group trained on an action video game recovered faster from the attentional blink than did the control group trained on a non-action video game ($*P < 0.05$, $**P < 0.01$).

eccentricity: 6 ms for targets presented at 10°; and 12 ms for targets present at either 20° or 30°.

For experiment 4—attentional blink (Fig. 4)—the stimuli and procedure described previously²⁰ were used except that eight lags instead of nine, and eight trials per condition instead of ten, were used. VGPs had almost significantly higher-second-target detection accuracy in the control condition (detecting the X in the absence of a first target, 95.6% to 87.9%, $P = 0.061$). Conservatively, all analyses were done on the difference between the variable of interest and the baseline detection condition to eliminate different baseline abilities as a potential confound in interpreting the depth of the attentional blink between populations.

For experiment 5—training (Fig. 5)—all subjects were tested on the useful-field-of-view, attentional-blink and enumeration tasks (in this order) before training; they then underwent training for 1 h per day for 10 days with one of two possible kinds of video game. After training they were retested on the same three tasks (in the same order). The experimental group trained on *Medal of Honor: Allied Assault* (Electronic Arts, 2002). This game simulates Second World War combat situations and was chosen to be similar to those played by our VGPs. It has a relatively simple interface, uses first-person point of view and requires effective monitoring of the whole visual field (extent from fixation about 13° (height) × 16° (width)). Subjects played the game straight through for the first 8 days. On days 9 and 10, they returned to the beginning of the game to quantitatively measure their improvement (comparing performance over mission 1 during their first and second playings). In all available measures of improvement taken (accuracy and deaths to complete mission), subjects improved their performance after training (accuracy, 17% improvement; deaths to complete, 42% improvement).

For the control group ($n = 8$), *Tetris* (Original Tetris, AcademySoft Elorgwhich, 1987; Tengen, 1988) was displayed to cover the entire extent of the screen. Because *Tetris* adds graphics on the side of the screen, however, the effective game area extended 13° (height) × 9° (width) from fixation. This game was selected to control for the effect of improved visuo-motor coordination, while putting few demands on the processing of several objects at once. Accordingly, the version of *Tetris* on which subjects were trained had the preview block option turned off. In the two measures of improvement available (high score, maximum level reached), all subjects improved after training (high score, 71% improvement; mean level improvement, 67% improvement).

Received 4 December 2002; accepted 8 April 2003; doi:10.1038/nature01647.

- Sagi, D. & Tanne, D. Perceptual learning: learning to see. *Curr. Opin. Neurobiol.* **4**, 195–199 (1994).
- Ball, K. & Sekuler, R. A specific and enduring improvement in visual motion discrimination. *Science* **218**, 697–698 (1982).
- Fiorentini, A. & Berardi, N. Perceptual learning specific for orientation and spatial frequency. *Nature* **287**, 43–44 (1980).
- Karni, A. & Sagi, D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natl Acad. Sci. USA* **88**, 4966–4970 (1991).
- Poggio, T., Fahle, M. & Edelman, S. Fast perceptual learning in visual hyperacuity. *Science* **256**, 1018–1021 (1992).
- Fahle, M. & Edelman, S. Long-term learning in vernier acuity: effects of stimulus orientation, range, and of feedback. *Vision Res.* **33**, 397–412 (1993).
- Ramachandran, V. S. & Braddick, O. Orientation-specific learning in stereopsis. *Perception* **2**, 371–376 (1973).
- Ahissar, M., Laiwand, R., Kozminsky, G. & Hochstein, S. Learning pop-out detection: building representations for conflicting target-distractor relationship. *Vision Res.* **38**, 3095–3107 (1998).
- Ahissar, M. & Hochstein, S. The spread of attention and learning in feature search: effects of target distribution and task difficulty. *Vision Res.* **40**, 1349–1364 (2000).
- Sirenteau, R. & Regina, R. Perceptual learning in visual search generalizes over task locations and eyes. *Vision Res.* **40**, 2925–2949 (2000).
- Eriksen, B. A. & Eriksen, C. W. Effects of noise letters upon the identification of a target letter in nonsearch task. *Percept. Psychophys.* **16**, 143–149 (1974).
- Lavie, N. & Cox, S. On the efficiency of visual selective attention: efficient visual search leads to inefficient distractor rejection. *Psychol. Sci.* **8**, 395–398 (1997).
- Kaufman, E., Lord, M., Reese, T. & Volkman, J. The discrimination of visual number. *Am. J. Psychol.* **62**, 498–525 (1949).
- Trick, L. M. & Pylyshyn, Z. W. What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *J. Exp. Psychol. Human Percept. Perform.* **19**, 331–351 (1993).
- Trick, L. M. & Pylyshyn, Z. W. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* **101**, 80–102 (1994).
- Tuholski, S. W., Engle, R. W. & Baylis, G. C. Individual differences in working memory capacity and enumeration. *Mem. Cogn.* **29**, 484–492 (2001).
- Ball, K., Beard, B. L., Roenker, D. L., Miller, R. L. & Griggs, D. S. Age and visual search: expanding the useful field of view. *J. Optic. Soc. Am.* **5**, 2210–2219 (1988).
- Myers, R. S., Ball, K. K., Kalina, T. D., Roth, D. L. & Goode, K. T. Relation of useful field of view and other screening tests to on-road driving performance. *Percept. Mot. Skills* **91**, 279–290 (2000).
- Ball, K. K., Owsley, C., Sloane, M., Roenker, D. & Bruni, J. Visual attention problems as a predictor of vehicle crashes in older drivers. *Invest. Ophthalmol.* **34**, 3110–3123 (1993).
- Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Human Percept. Perform.* **18**, 849–860 (1992).
- Chun, M. M. & Potter, M. C. A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Human Percept. Perform.* **21**, 109–127 (1995).
- Broadbent, D. & Broadbent, M. From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept. Psychophys.* **42**, 105–113 (1987).
- Arnell, K. & Jolicoeur, P. The attentional blink across stimulus modalities: evidence for central processing limitation. *J. Exp. Psychol. Human Percept. Perform.* **25**, 630–648 (1999).
- Potter, M. C., Chun, M. M., Banks, B. S. & Muckenhoupt, M. Two attentional deficits in serial target

search: the visual attentional blink and an amodal task-switch deficit. *J. Exp. Psychol. Human Percept. Perform.* **24**, 979–992 (1998).

- Prokisch, J. & Bavelier, D. Changes in the spatial distribution of visual attention after early deafness. *J. Cogn. Neurosci.* **14**, 1–5 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Monacelli, D. McColgin, J. Cohen and K. Schneider for help with subjects and software support; and R. Aslin, A. Pouget and D. Knill for feedback on the manuscript. This research was supported by the NIH and the James S. McDonnell-Pew Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (daphne@cvs.rochester.edu).

Mitochondrial membrane remodelling regulated by a conserved rhomboid protease

G. Angus McQuibban, Saroj Saurya & Matthew Freeman

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Rhomboid proteins are intramembrane serine proteases that activate epidermal growth factor receptor (EGFR) signalling in *Drosophila*¹. Rhomboids are conserved throughout evolution^{2–5}, and even in eukaryotes their existence in species with no EGFRs implies that they must have additional roles. Here we report that *Saccharomyces cerevisiae* has two rhomboids, which we have named Rbd1p and Rbd2p. *RBD1* deletion results in a respiratory defect; consistent with this, Rbd1p is localized in the inner mitochondrial membrane and mutant cells have disrupted mitochondria. We have identified two substrates of Rbd1p: cytochrome *c* peroxidase (Ccp1p); and a dynamin-like GTPase (Mgm1p), which is involved in mitochondrial membrane fusion^{6–10}. Rbd1p mutants are indistinguishable from Mgm1p mutants, indicating that Mgm1p is a key substrate of Rbd1p and explaining the *rbd1Δ* mitochondrial phenotype. Our data indicate that mitochondrial membrane remodelling is regulated by cleavage of Mgm1p and show that intramembrane proteolysis by rhomboids controls cellular processes other than signalling. In addition, mitochondrial rhomboids are conserved throughout eukaryotes and the mammalian homologue, PARL¹¹, rescues the yeast mutant, suggesting that these proteins represent a functionally conserved subclass of rhomboid proteases.

Despite their widespread conservation, the only known function of eukaryotic rhomboid proteases is the activation of EGFR signalling in *Drosophila*^{12–15}. We therefore examined their function in *S. cerevisiae*, which has no receptor tyrosine kinases but has two genes encoding rhomboids, which we have named *RBD1* (*YGR101w*) and *RBD2* (*YPL246c*). Deletion of *RBD1* (*Δrbd1*) caused cells to grow slowly (Fig. 1a), whereas deletion of *RBD2* had no obvious phenotype and did not enhance the phenotype of *rbd1Δ* cells (data not shown). The *rbd1Δ* phenotype was rescued by plasmid-borne expression of *RBD1* (Fig. 1a), confirming that the slow growth was indeed caused by loss of *RBD1*.

Although previously unnoticed, the Rbd1p sequence contains a signature motif for mitochondrial localization, as predicted by MitoProt¹⁶ and other algorithms. To test this prediction, we used homologous recombination to replace the wild-type gene with a fully functional gene fused at its carboxy terminus to green

fluorescent protein (GFP; see Supplementary Fig. S1). Rbd1p–GFP colocalized precisely with antibodies against yeast porin, a mitochondrial protein, showing that Rbd1p is restricted to the mitochondria (Fig. 1b).

Rbd1p is an integral membrane protein with six predicted

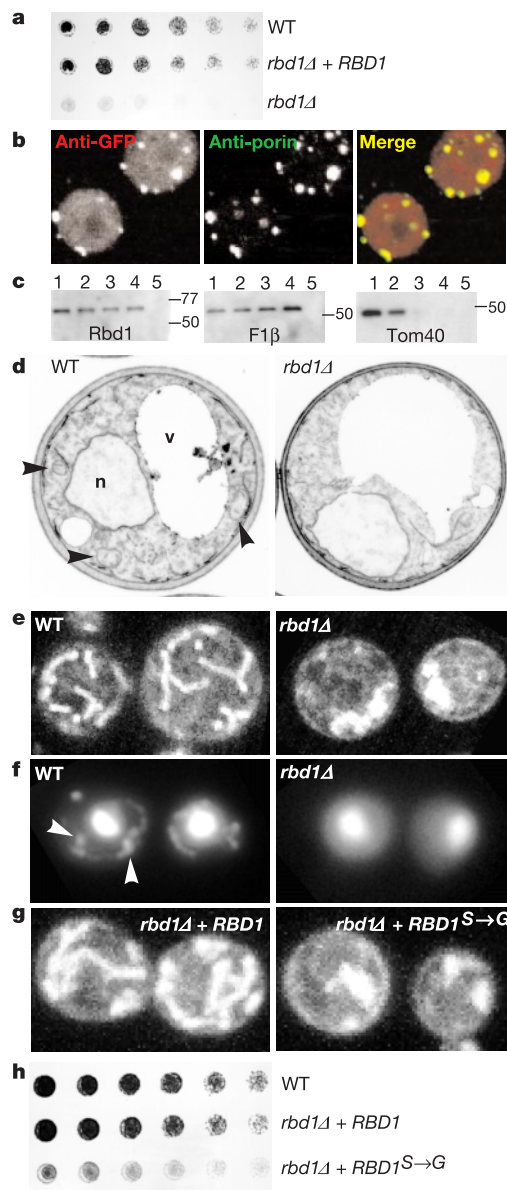


Figure 1 Characterization of Rbd1p. **a**, *rbd1Δ* cells grow more slowly than wild type. This growth defect is rescued by plasmid-borne expression of *RBD1*. **b**, Optical sections through yeast spheroplasts indicate the mitochondrial localization of Rbd1p–GFP (red), as shown by its colocalization with yeast mitochondrial porin (green). **c**, Rbd1p has an inner mitochondrial membrane location. Enriched mitochondrial fractions were treated with increasing amounts of proteinase K (lanes 1–4, 0, 10, 50 and 100 $\mu\text{g ml}^{-1}$; lanes 5, 100 $\mu\text{g ml}^{-1}$ plus 0.1% Triton X-100). Rbd1p was protected from degradation except in the presence of detergent. Compare the markers F1 β and Tom40p, which are respectively inner and outer membrane proteins. **d**, Electron microscopy shows a loss of normal mitochondria in *rbd1Δ* cells. The nucleus (n) and vacuole (v) are labelled, and normal mitochondria are indicated with black arrowheads. **e**, Vital staining of wild-type and *rbd1Δ* cells with Mitotracker Green FM. **f**, DAPI staining of wild-type and *rbd1Δ* cells, showing loss of nucleoids. **g**, Mitochondrial morphology defects are rescued by a plasmid-borne wild-type copy of *RBD1* (91% wild-type mitochondria, 9% fragmented/aggregated, $n = 100$) but not mutant *RBD1* (0% wild-type mitochondria, 100% fragmented/aggregated, $n = 100$). **h**, Growth of *rbd1Δ* is also fully restored by wild-type but not mutant *RBD1*.

transmembrane domains (TMDs), and we localized it to the mitochondrial inner membrane by monitoring the protease digestion of intact mitochondria (Fig. 1c). This clearly distinguishes Rbd1p from previously analysed eukaryotic rhomboids, all of which have been found to be located in the secretory pathway or on the plasma membrane¹⁷. Combined with the slow growth phenotype, the mitochondrial location of Rbd1p suggested that *rbd1Δ* cells might have a mitochondrial defect. Consistent with this, the *rbd1Δ* strain failed to grow on the non-fermentable carbon source glycerol (Fig. 2b), suggesting that it was deficient in respiration.

To test directly the possibility that loss of Rbd1p caused mitochondrial defects, we examined mutant cells by electron microscopy and found that they lacked wild-type mitochondria (Fig. 1d). The cells appeared otherwise normal, and other intracellular structures were indistinguishable from wild-type controls. We next used a

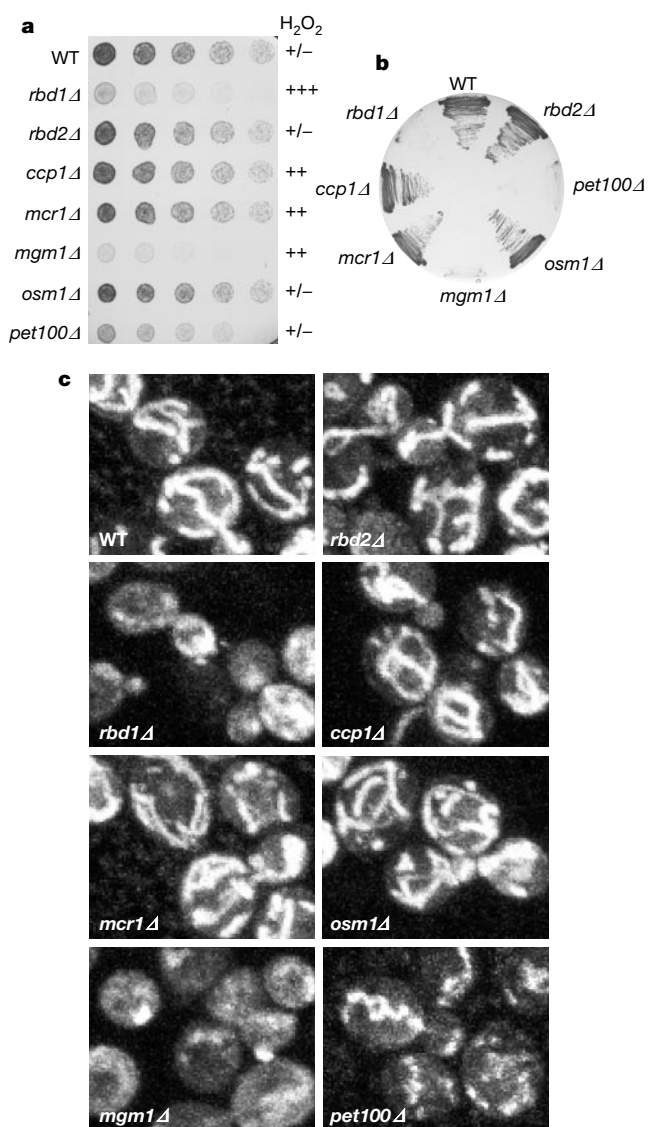
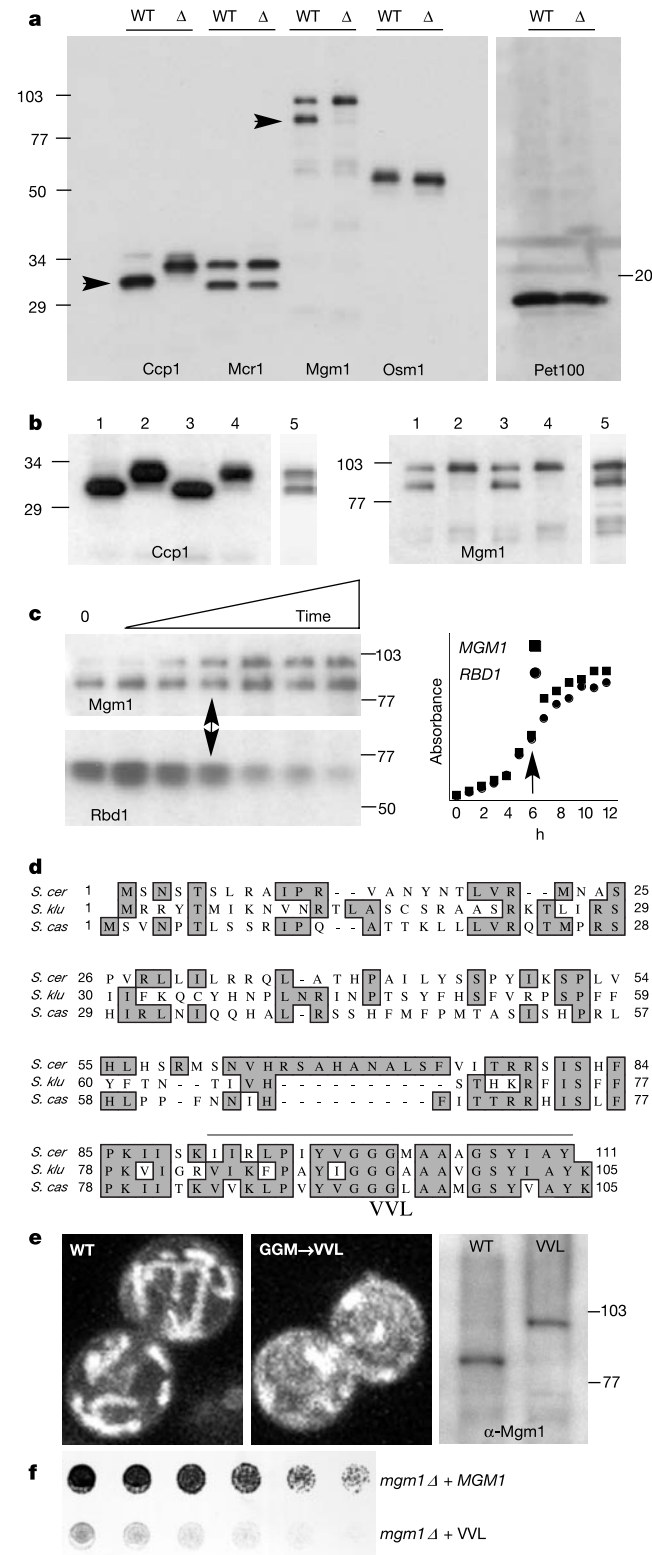


Figure 2 Phenotypes of candidate Rbd1p substrates. **a**, Growth of five strains with deletions of potential substrates is compared with the growth of *rbd1Δ*. Only *mgm1Δ* and *pet100Δ* show growth defects similar to *rbd1Δ*. The relative sensitivity of each strain to H_2O_2 is shown on the right. *rbd1Δ*, *ccp1Δ*, *mcr1Δ* and *mgm1Δ* are sensitive to H_2O_2 . **b**, Growth on glycerol-containing medium shows that *rbd1Δ*, *mgm1Δ* and *pet100Δ* are respiration-deficient. **c**, Of the candidates tested, the mitochondrial morphology (as detected by Mitotracker Green FM) of only *mgm1Δ* resembles that of *rbd1Δ*. In all tests, *rbd2Δ* was included as a control for specificity of the phenotypes.

mitochondrion-specific dye to examine living cells. Mitochondria from wild-type yeast cells in log-phase growth generally appear as tubular structures around the cell cortex. By contrast, the mitochondria of *rbd1Δ* cells appeared as small fragments and aggregated masses throughout the cell (Fig. 1e). In further support of mitochondrial defects, 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) staining detected a loss of nucleoid structures, which represent mitochondrial genomes (Fig. 1f).



The requirement for a rhomboid in the maintenance of mitochondrial morphology and genome maintenance was unexpected and suggested two possibilities. Either intramembrane proteolytic activity similar to that which activates intercellular signalling in other cases is used in a different context, or some other uncharacterized feature of the rhomboid protein is responsible for its mitochondrial function. To distinguish between these possibilities, we tested whether wild-type mitochondrial morphology and growth rates depended on the catalytic activity of Rbd1p. Expression of wild-type *RBD1* rescued the *rbd1Δ* cells, but a mutant form of the gene, in which the catalytic serine is replaced by glycine¹, failed to rescue the phenotype (Fig. 1g, h). This indicates that the intramembrane proteolytic activity of Rbd1p is required for its mitochondrial function.

To investigate further this requirement for a mitochondrial protease, we identified potential Rbd1p substrates by selecting proteins from the yeast genome that met the following criteria: a mitochondrial localization, the presence of a single predicted TMD and experimental evidence that the protein is soluble, which together suggest that the protein may undergo proteolytic cleavage. Five characterized proteins fulfilled these criteria: Ccp1p, a cytochrome *c* peroxidase¹⁸; Mcr1p, an NADH-cytochrome *b*₅ reductase¹⁹; Mgm1p, a dynamin-like GTPase⁶; Osm1p, a fumarate reductase²⁰; and Pet100p, a chaperone required by cytochrome *c* oxidase²¹. We examined deletion strains of each of these candidate substrates in four assays of mitochondrial function: overall growth rate, peroxide sensitivity, growth on glycerol and mitochondrial morphology (Fig. 2a–c). Only *mgm1Δ* was indistinguishable in behaviour from *rbd1Δ*.

We examined directly whether any of the candidate substrates undergo Rbd1p-dependent processing *in vivo*, by replacing the wild-type genome copy of the gene with a C-terminal haemagglutinin A (HA)-tagged copy, both in the wild-type and *rbd1Δ* strains. Whereas Mcr1p, Osm1p and Pet100p were unaffected by the loss of Rbd1p, Ccp1p and Mgm1p were cleaved in an Rbd1p-dependent way (Fig. 3a). In late log phase, all of Ccp1p and about 50% of Mgm1p were processed in wild-type cells. Both proteins, however, were uncleaved in *rbd1Δ* cells; this requirement for Rbd1p in Ccp1p processing confirms the results of another study²².

Proteolytic cleavage was rescued in the *rbd1Δ* strain by a wild-type copy of *RBD1* but not by a catalytically inactive form

Figure 3 Cleavage of Ccp1p and Mgm1p depends on Rbd1p *in vivo*. **a**, Western blot analysis of cell lysates from cultures at late log phase shows that Ccp1p and Mgm1p are each processed to a faster-migrating species (arrowheads) in wild type but not in *rbd1Δ*. By contrast, Mcr1p, Osm1p and Pet100p are unaffected by the loss of Rbd1p. **b**, Cleavage of both Ccp1p and Mgm1p is restored in *rbd1Δ* strains by plasmid-borne expression of *RBD1* (lanes 3) but not a catalytic mutant form of *RBD1* (lanes 4). Plasmid-borne expression of the human homologue of *RBD1* (hPARI) also restores cleavage of both Ccp1p and Mgm1p (lanes 5). Lanes 1, cell lysates from wild type; lanes 2, cell lysates from *rbd1Δ*. **c**, Western blot analysis of Mgm1p and Rbd1p (using antibodies against HA and GFP, respectively) in cell lysates from wild-type cells as they exit logarithmic growth. Time points from left to right are 0, 2, 4, 6, 8, 10 and 12 h, starting at an absorbance at 600 nm of 0.2. Growth curves indicate that cells start to exit logarithmic phase at ~6 h (arrowheads). **d**, Alignment of the N termini of Mgm1p from *S. cerevisiae* and from two other yeast species, *Saccharomyces castellii* and *Saccharomyces kluyveri* (P. Clifton and M. Johnston, personal communication). Horizontal line indicates the predicted TMD; VVL indicates the position of the TMD mutant used in **e**. **e**, Mitochondrial morphology defects are rescued by plasmid-borne wild-type copy of *MGM1* (83% wild-type mitochondria, 17% fragmented/aggregated, *n* = 100) but not a mutant copy of *MGM1*, in which TMD residues 101–103 have been replaced with VVL (0% wild-type mitochondria, 100% fragmented/aggregated, *n* = 100). Western blot of lysates of cultures in logarithmic phase shows that the VVL mutant Mgm1p is not efficiently cleaved. **f**, The slow growth phenotype is rescued by a plasmid-borne copy of wild-type *MGM1* but not a copy of the VVL *MGM1* mutant.

(Fig. 3b), showing that, as with the mitochondrial phenotype, the processing of both Ccp1p and Mgm1p requires the intramembrane serine protease activity of Rbd1p. The normal processing of Mcr1p in *rbd1Δ* cells (which depends on a different mitochondrial protease¹⁹) indicates that Rbd1p is not generally required for cleavage of mitochondrial proteins. These data explain the identical phenotype of *mgm1Δ* and *rbd1Δ* and, coupled with the very weak phenotype of *ccp1Δ*, strongly suggest that Mgm1p is the primary substrate of the proteolytic function of Rbd1p in maintaining the integrity of the mitochondrial membrane. Our results are also consistent with data implying that Mgm1p regulates mitochondrial membrane fusion^{7,8,10}.

Our data indicate that, in order to function in membrane fusion, Mgm1p needs to be activated by Rbd1p-catalysed intramembrane cleavage. Because mitochondrial morphology and requirements change significantly as cells move from exponential growth to stationary phase, we examined the amounts of Rbd1p protein and cleavage of Mgm1p during this transition period. As cells left logarithmic-phase growth, Rbd1p was downregulated and there was a simultaneous reduction of Mgm1p cleavage, from 95% to about 50% (Fig. 3c). Notably, the overall level of Mgm1p did not decrease over the time course of the experiment, implying that the reduction of Rbd1p did not simply reflect a general loss of mitochondrial protein. This suggests that expression of Rbd1p is a physiologically significant regulator of mitochondrial remodelling.

Analysis of Mgm1p sequences from divergent yeasts (Fig. 3d) identifies a highly conserved region that is predicted to form a TMD and represents a potential cleavage site for Rbd1p. Notably, this region has sequence characteristics (glycine and alanine residues) that suggest that it might be a rhomboid substrate²³. To test this prediction, we altered amino acids 101–103 of Mgm1p (GGM to VVL) and tested the ability of this TMD mutation to rescue the *mgm1Δ* strain. This mutant did not complement the mitochondrial morphology or growth of the *mgm1Δ* strain and was uncleaved in cell extracts (Fig. 3e, f). These data support the hypothesis that Rbd1p cleaves Mgm1p in this proposed TMD.

The rhomboid protease family is widely conserved^{2–5} and, using the MitoProt algorithm for prediction of mitochondrial targeting domains, we found that mitochondrial rhomboids occur throughout the eukaryotes (Fig. 4a): in *Drosophila*, it is Rhomboid-7; and in mammals, PARL (presenilin-associated rhomboid-like; accession number AAG28519)¹¹. Neither of these has an assigned function. The prediction of mitochondrial location was validated by expressing mouse PARL in COS cells, where it was localized exclusively to mitochondria (Fig. 4b). We also examined the possibility that the function, as well as the location, of mitochondrial rhomboids is conserved, by testing whether human PARL could rescue the *rbd1Δ* mutation. The expression of PARL in *rbd1Δ* cells restored Ccp1p and Mgm1p processing (Fig. 3b, lanes 5) and also rescued growth rate and mitochondrial morphology (data not shown). This suggests that the mitochondrial function of Rbd1p might be conserved in mammals. Our combined data identify a subclass of rhomboids that control mitochondrial membrane dynamics in yeast and provide evidence that their function and specificity is conserved in mammals.

Mitochondria are dynamic organelles, frequently undergoing marked remodelling^{24–26}. A balance of fusion and fission events is thought to regulate this process and Mgm1p is one of a trio of dynamin-like GTPases, conserved from yeast to mammals, that control these membrane dynamics^{7,8}. Mgm1p functions in the intermembrane space and regulates membrane fusion by interacting with the outer membrane proteins Fzo1p and Ugo1p¹⁰. Our data imply that Mgm1p is initially an integral inner membrane protein (Fig. 3d, e, and Supplementary Fig. S2); consistent with this, full-length and cleaved forms of Mgm1p show differential membrane association⁷. Our results also imply that intramem-

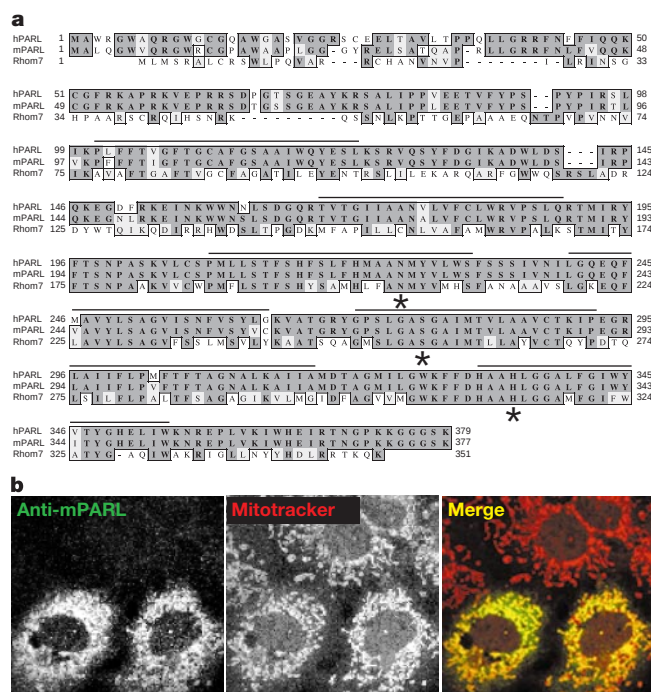


Figure 4 Mitochondrial rhomboids are conserved in higher eukaryotes. **a**, Sequence alignment of the human (hPARL), mouse (mPARL) and *Drosophila* (Rhomboid-7) mitochondrial rhomboids. Bars above the sequence indicate the predicted TMDs (using TMHMM v.2.0) and the three asterisks below the sequence indicate the putative catalytic triad residues¹. **b**, Expression of mouse PARL in COS cells shows colocalization with the mitochondrion-specific dye Mitotracker.

brane cleavage of the TMD of this GTPase may be an essential activation step for its function. This is a previously unrecognized mode of regulation for dynamin-like proteins and suggests that membrane tethering may be incompatible with dynamin-like membrane remodelling activity. Notably, heterozygosity for the mammalian homologue of Mgm1p, OPA1 (ref. 27), is the cause of the most common form of childhood-onset blindness, dominant optic atrophy²⁸. It also has a predicted TMD and regulates mitochondrial membrane dynamics²⁸. Because human PARL can complement *rbd1Δ*, it is an intriguing possibility that the human mitochondrial rhomboid might also be involved in regulating the activity of OPA1. □

Methods

Construction of yeast strains

All genetic manipulation of yeast was done as described²⁹. Disruptions of the *S. cerevisiae* open reading frames *YGR101w* (also called Pcp1p²³) and *YPL246c* were made in the α -303 yeast strain. We tagged candidate substrates at the C terminus with the influenza HA sequence by homologous recombination. Integration was confirmed by polymerase chain reaction (PCR) analysis of mutant colonies. Mutant strains of candidate Rbd1p substrates and *RBD1* and *RBD2* were purchased from EUROSCARF. For rescue experiments, the *RBD1* open reading frame plus 200 nucleotides at the 5' and 3' ends was amplified by PCR from yeast genomic DNA and cloned into the yeast vector pRS315 (ATCC); the *MGM1* gene in pRS313 was a gift from M. Yaffe.

Mutant strains carrying test plasmids were created by either plasmid shuffle (*RBD1* rescue) or by tetrad dissection of transformed diploids (*RBD1* and *MGM1* rescue). We carried out site-directed mutagenesis of the catalytic serine residue of *RBD1* to glycine and the transmembrane region of *MGM1* using the Quick-Change kit (Stratagene). The complementary DNA of human PARL (obtained from IMAGE consortium) was cloned into a derivative of pRS315 with the Pho5 promoter (a gift from J. Whyte). Antisera against Mgm1p were a gift from M. Yaffe. Antisera against Tom40p were a gift from B. Westermann. Antisera against F1 β were a gift from R. Jensen.

Histology

Electron microscopy images were taken after staining yeast cells with permanganate and then staining sections with uranyl acetate. Live yeast cells were fluorescently labelled with Mitotracker Green FM according to the manufacturer's instructions (Molecular Probes).

We carried out antibody staining on fixed spheroplasts from yeast cells, which were labelled with monoclonal antibodies against yeast porin (Molecular Probes) and a rabbit polyclonal antiserum generated against recombinant GFP¹². Nucleoid bodies were stained with DAPI by standard protocols²⁹.

Protease protection

Mitochondrion-enriched fractions were prepared as described³⁰ from wild-type, *RBD1:GFP* and *MGM1:HA* yeast strains. We carried out protease protection assays as described⁸ and analysed samples by western blot using the indicated antibodies.

Phenotypic analysis

Growth was tested on plates containing 2% glucose. Respiration competence was tested by growth on plates containing 2% glycerol. We assessed peroxide sensitivity by measuring the zone of cell death around a 5-mm disk containing 5 µl of 30% hydrogen peroxide. All growth assays were done at 30 °C.

Cell culture

We cultured COS cells under standard conditions. The mouse PARL cDNA (a gift from B. De Strooper) was inserted into the expression vector pSGF5 (Stratagene) and was transfected using Lipofectamine 2000, according to the manufacturer's instructions (Invitrogen). Forty-eight hours after transfection, cells were stained with Mitotracker (Molecular Probes), fixed with 4% formaldehyde and counterstained with a monoclonal antibody against the C-terminal HA tag as described¹².

Received 26 February; accepted 1 April 2003; doi:10.1038/nature01633.

- Urban, S., Lee, J. R. & Freeman, M. *Drosophila* Rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**, 173–182 (2001).
- Wasserman, J. D., Urban, S. & Freeman, M. A family of rhomboid-like genes: *Drosophila* rhomboid-1 and rhomboid-3 cooperate to activate EGF receptor signalling. *Genes Dev.* **14**, 1651–1663 (2000).
- Pascall, J. C. & Brown, K. D. Characterization of a mammalian cDNA encoding a protein with high sequence similarity to the *Drosophila* regulatory protein Rhomboid. *FEBS Lett.* **429**, 337–340 (1998).
- Urban, S., Schlieper, D. & Freeman, M. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr. Biol.* **12**, 1507–1512 (2002).
- Koonin, E. V. *et al.* The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* **4**, R19 www.genomebiology.com (2003).
- Jones, B. A. & Fangman, W. L. Mitochondrial DNA maintenance in yeast requires a protein containing a region related to the GTP-binding domain of dynamin. *Genes Dev.* **6**, 380–389 (1992).
- Shepard, K. A. & Yaffe, M. P. The yeast dynamin-like protein, Mgm1p, functions on the mitochondrial outer membrane to mediate mitochondrial inheritance. *J. Cell Biol.* **144**, 711–720 (1999).
- Wong, E. D. *et al.* The dynamin-related GTPase, Mgm1p, is an intermembrane space protein required for maintenance of fusion competent mitochondria. *J. Cell Biol.* **151**, 341–352 (2000).
- Sato, M., Hamamoto, T., Seo, N., Kagawa, Y. & Endo, H. Differential sublocalization of the dynamin-related protein OPA1 isoforms in mitochondria. *Biochem. Biophys. Res. Commun.* **300**, 482–493 (2003).
- Wong, E. D. *et al.* The intramitochondrial dynamin-related GTPase, Mgm1p, is a component of a protein complex that mediates mitochondrial fusion. *J. Cell Biol.* **160**, 303–311 (2003).
- Pellegrini, L. *et al.* PAMP and PARL, two novel putative metalloproteases interacting with the COOH-terminus of Presenilin-1 and -2. *J. Alzheimers Dis.* **3**, 181–190 (2001).
- Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. Regulated intracellular ligand transport and proteolysis controls EGF signal activation in *Drosophila*. *Cell* **107**, 161–171 (2001).
- Tsruya, R. *et al.* Intracellular trafficking by Star regulates cleavage of the *Drosophila* EGF receptor ligand Spitz. *Genes Dev.* **16**, 222–234 (2002).
- Ghigliione, C. *et al.* Mechanism of activation of the *Drosophila* EGF receptor by the TGF α ligand Gurken during oogenesis. *Development* **129**, 175–186 (2002).
- Klämbt, C. EGF receptor signalling: roles of star and rhomboid revealed. *Curr. Biol.* **12**, R21–R23 (2002).
- Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
- Urban, S., Lee, J. R. & Freeman, M. A family of Rhomboid intramembrane proteases activates all *Drosophila* membrane-tethered EGF ligands. *EMBO J.* **21**, 4277–4286 (2002).
- Kaput, J., Goltz, S. & Blobel, G. Nucleotide sequence of the yeast nuclear gene for cytochrome c peroxidase precursor. Functional implications of the pre sequence for protein transport into mitochondria. *J. Biol. Chem.* **257**, 15054–15058 (1982).
- Hahne, K., Haucke, V., Ramage, L. & Schatz, G. Incomplete arrest in the outer membrane sorts NADH-cytochrome *b₅* reductase to two different submitochondrial compartments. *Cell* **79**, 829–839 (1994).
- Muratsubaki, H. & Enomoto, K. One of the fumarate reductase isoenzymes from *Saccharomyces cerevisiae* is encoded by the *OSM1* gene. *Arch. Biochem. Biophys.* **352**, 175–181 (1998).
- Church, C., Chapon, C. & Poyton, R. O. Cloning and characterization of PET100, a gene required for the assembly of yeast cytochrome c oxidase. *J. Biol. Chem.* **271**, 18499–18507 (1996).
- Esser, K., Tursun, B., Ingenhoven, M., Michaelis, G. & Pratz, E. A novel two-step mechanism for removal of a mitochondrial signal sequence involves the mAAA complex and the putative rhomboid protease Pcp1. *J. Mol. Biol.* **323**, 835–843 (2002).
- Urban, S. & Freeman, M. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell* (in the press).
- Hermann, G. J. & Shaw, J. M. Mitochondrial dynamics in yeast. *Annu. Rev. Cell Dev. Biol.* **14**, 265–303 (1998).
- Yaffe, M. P. Dynamic mitochondria. *Nature Cell Biol.* **1**, E149–E150 (1999).
- Griparic, L. & van der Bliek, A. M. The many shapes of mitochondrial membranes. *Traffic* **2**, 235–244 (2001).

- Alexander, C. *et al.* OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nature Genet.* **26**, 211–215 (2000).
- Delettre, C., Lenaers, G., Pelloquin, L., Belenguer, P. & Hamel, C. P. OPA1 (Kjer type) dominant optic atrophy: a novel mitochondrial disease. *Mol. Genet. Metab.* **75**, 97–107 (2002).
- Guthrie, C. & Fink, G. R. (eds) *Guide to Yeast Genetics* (Academic, San Diego, 2002).
- Daum, G., Bohni, P. C. & Schatz, G. Import of proteins into mitochondria. Cytochrome *b₂* and cytochrome c peroxidase are located in the intermembrane space of yeast mitochondria. *J. Biol. Chem.* **257**, 13028–13033 (1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Pelham, S. Munro and J. Whyte for help and support; B. De Strooper for discussions; members of our laboratory for advice; M. Yaffe, R. Jensen and B. Westermann for antibodies; and P. Clifton and M. Johnston for providing yeast sequences before publication. G.A.M. is supported by an EMBO long-term fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F. (MF1@mrc-lmb.cam.ac.uk).

Role of ERas in promoting tumour-like properties in mouse embryonic stem cells

Kazutoshi Takahashi, Kaoru Mitsui & Shinya Yamanaka

Laboratory of Animal Molecular Technology, Research and Education Center for Genetic Information, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan

Embryonic stem (ES) cells are pluripotent cells derived from early mammalian embryos^{1,2}. Their immortality and rapid growth make them attractive sources for stem cell therapies³; however, they produce tumours (teratomas) when transplanted, which could preclude their therapeutic usage⁴. Why ES cells, which lack chromosomal abnormalities, possess tumour-like properties is largely unknown. Here we show that mouse ES cells specifically express a Ras-like gene, which we have named *ERas*. We show that human *HRas*, which is a recognized pseudogene, does not contain reported base substitutions and instead encodes the human orthologue of ERas. This protein contains amino-acid residues identical to those present in active mutants of Ras⁵ and causes oncogenic transformation in NIH 3T3 cells. ERas interacts with phosphatidylinositol-3-OH kinase⁶ but not with Raf^{7,8}. *ERas*-null ES cells maintain pluripotency but show significantly reduced growth and tumorigenicity, which are rescued by expression of *ERas* complementary DNA or by activated phosphatidylinositol-3-OH kinase. We conclude that the transforming oncogene *ERas* is important in the tumour-like growth properties of ES cells.

To understand mechanisms underlying the pluripotency of ES cells and their propensity to form tumours *in vivo*, we searched for genes that were expressed specifically in murine ES cells by digital differential display. Unigene cluster Mm.249524 was found exclusively in libraries from ES cells. We obtained a full-length cDNA of this gene by the rapid amplification of cDNA ends (RACE). Basic local alignment search tool (BLAST) analysis against mouse genomic databases showed that the gene is located on the X chromosome and contains two exons. The cDNA encodes a protein of 227 amino acids with 43%, 46% and 47% identity to HRas, KRas and NRas, respectively. Five domains essential for small G proteins⁹ are highly conserved (Fig. 1a). It contains a CAAX motif^{10,11} and is located at cytoplasmic membrane (Fig. 1b). These findings indicate that the

protein, which we designated ERas, is a previously unknown member of the Ras protein family.

By searching human genomic databases, we found that the *ERas* gene was most similar to human *HRas* (*Ha-Ras2*), which is also located on the X chromosome and is a recognized processed pseudogene with several nonsense and deletion mutations¹². We re-determined the nucleotide sequence of *HRas* and found that the previously reported mutations did not exist. *HRas* has a single open reading frame encoding a polypeptide with 76% identity to mouse ERas (Fig. 1a). Because of the sequence similarity and the common chromosomal localization, we concluded that *HRas* is the human orthologue of *ERas*.

Northern blot analyses detected a single 1.2-kb *ERas* transcript in undifferentiated RF8 mouse ES cells (ref. 13 and Fig. 1c). ERas was also expressed in three other undifferentiated mouse ES cell lines (Fig. 1d): J1 (ref. 14), CGR8 (ref. 15) and MG1.19 (ref. 16). In contrast, *ERas* was not detected in differentiated ES cells or in 12

somatic tissues from adult mouse (Fig. 1c). Therefore, the specific expression of *ERas* is a common property of ES cells.

Point mutations of several amino acids of HRas, including Gly 12, Ala 59 or Glu 63, render the protein constitutively active¹⁷. We found that mouse ERas has serine, serine and isoleucine (and human ERas has serine, alanine and asparagine) at the positions corresponding to Gly 12, Ala 59 and Glu 63 of HRas (Fig. 1a), suggesting that both were constitutively active. To explore this possibility, we analysed the association of GTP and GDP with the proteins (Fig. 2a). As expected, 95% of mouse and human ERas was in a GTP-bound form. In contrast, about 90% of wild-type HRas existed in a GDP-bound form, whereas about 80% of HRasV12 (a constitutively active mutant in which Gly 12 is mutated to valine) was in a GTP-bound form. These data show that ERas is indeed constitutively active.

To examine whether ERas has the ability to transform cells, we used an efficient retroviral gene transfer system to generate a polyclonal NIH 3T3 cell population expressing ERas or HRasV12 (Supplementary Fig. S1). Both mouse and human ERas induced morphological changes indicative of transformation: high refractivity and spindle-like shape (Fig. 2b) and loss of contact inhibition (Fig. 2c). When we cultured 5,000 NIH 3T3 cells expressing mouse ERas, human ERas or HRasV12 in soft agar, we obtained similar numbers (992 ± 160 , 934 ± 49 and 989 ± 86) of colonies showing anchorage-independent growth (Fig. 2d).

By contrast, from control cells transfected with the parent vector or ERas inactivated by deletion of the CAAX motif (ERas- Δ C), we obtained a few (2 ± 1 and 1 ± 1) colonies that were much smaller than those from wild-type ERas and HRasV12. When transplanted into nude mice, cells expressing ERas effectively produced tumours (Fig. 2e). ERas was more potent than HRasV12 in promoting cell growth in culture (Fig. 2c, d) but produced smaller tumours (Fig. 2e). Their difference was more evident in mouse embryonic fibroblasts (MEFs): HRasV12 caused premature senescence (ref. 18 and Fig. 2f), whereas ERas significantly increased growth rate. These results show that ERas induces transformation as effectively as HRasV12, but with a different mode of action.

To study the functions of ERas in ES cells, we transfected either the sense or antisense cDNA of *ERas* into MG1.19 ES cells. When the sense cDNA was transfected, expression of ERas increased roughly fourfold (Fig. 3a) and cell growth was significantly enhanced (Fig. 3b). Transfection of ERas- Δ C did not produce such growth-promoting effects. In contrast, transfection of HRasV12 resulted in differentiation and growth retardation (Fig. 3b), as reported previously^{19,20}. After transfection of the antisense cDNA, ERas expression decreased to a fifth of the normal level (Fig. 3a) and cell growth was significantly repressed (Fig. 3b). These data suggest that ERas, unlike HRasV12, promotes the growth of ES cells.

To clarify further the *in vivo* functions of ERas, we disrupted the *ERas* gene by homologous recombination in another ES cell line, RF8. We obtained two clones in which the coding region was replaced with a fusion comprising the β -galactosidase and neomycin resistance genes (β -geo; Fig. 3c). As expected from its location on the X chromosome, the wild-type *ERas* allele was absent in both clones, as assessed by Southern blot analysis (Fig. 3d). Northern blot (Fig. 3e), western blot (Fig. 3f) and immunohistochemical assessment (Fig. 1b) confirmed the absence of ERas expression. In addition, we established two clones, one from each knockout clone, in which ERas expression was partially recovered by transfection with *ERas* cDNA (Fig. 3f).

ERas-null ES cells were normal in both morphology (data not shown) and expression of *Oct3/4* (Fig. 3e). Blastocyst injection of these cells resulted in germline transmission of the mutation. Mutant mice did not show gross abnormalities or infertility (data not shown). These results indicate that ERas is not required for pluripotency and normal mouse development. However, *ERas*-null cells grew significantly more slowly than wild-type cells (Fig. 3g).

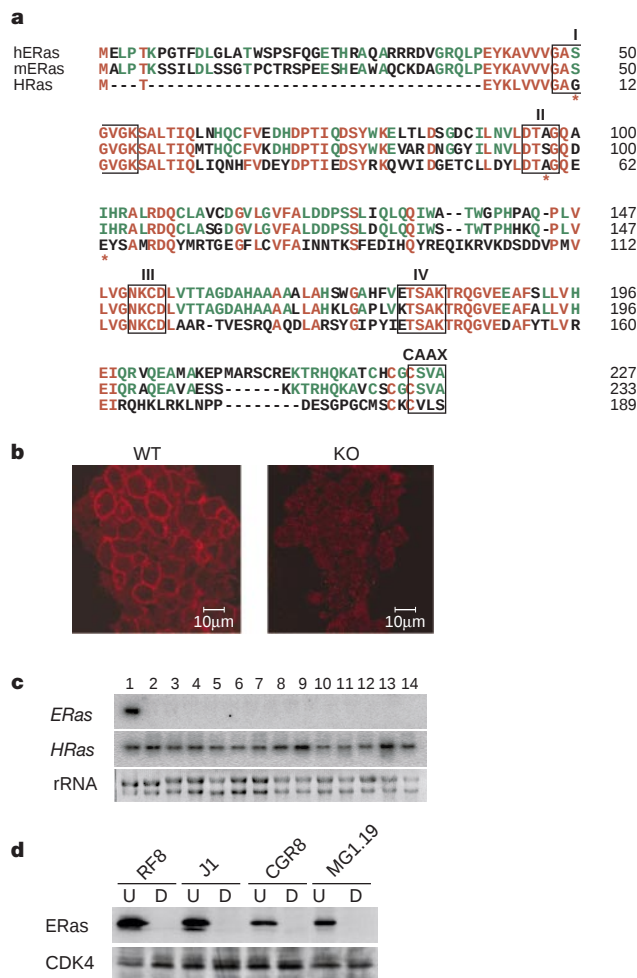


Figure 1 Identification of ERas. **a**, Amino acid sequences of murine and human ERas and HRas. Boxes indicate conserved domains and the CAAX motif. Asterisks indicate three amino acid residues that are often mutated in oncogenic HRas mutants. Amino acids conserved among all three proteins are red, and those conserved between human and mouse ERas are green. **b**, Immunostaining of ERas in wild-type (WT) and *ERas* knockout (KO) ES cells. **c**, Northern blot. Lane 1, undifferentiated ES cells; lane 2, retinoic-acid-treated ES cells; lane 3, testis; lane 4, lung; lane 5, heart; lane 6, liver; lane 7, stomach; lane 8, kidney; lane 9, brain; lane 10, spleen; lane 11, thymus; lane 12, small intestine; lane 13, skin; lane 14, muscle. **d**, Western blot analysis of ERas expression in four ES cell lines. U, undifferentiated; D, differentiated.

The slow growth phenotype was more evident when ES cells were grown without feeder cells. Wild-type ES cells expanded 6×10^5 times under these suboptimal conditions over 16 d, whereas the *ERas*-null cells expanded only 1×10^3 times. When transplanted into nude mice, *ERas*-deficient cells produced markedly smaller tumours than did wild-type cells (Fig. 3h). Re-expression of *ERas* led to a partial recovery of growth and tumorigenicity (Fig. 3g, h). These results show the important role of ERas in the tumour-like growth properties of ES cells.

We thought that the functional differences between ERas and HRasV12 might be due to different effector preferences. To test this hypothesis, we determined whether ERas could activate Raf^{7,8}. Co-immunoprecipitation experiments showed that HRasV12, but not ERas, associates with Raf1 (Fig. 4a) and B-Raf (Supplementary Fig. S2a). In reporter assays, AP1 enhancer activity was activated by HRasV12 and repressed by the dominant-negative mutant HRasN17 (Supplementary Fig. S2b). In contrast, ERas did not affect AP1 activity. These data indicate that ERas cannot bind to Raf or activate the mitogen-activated protein kinase (MAPK) cascade.

We next examined whether ERas binds to phosphatidylinositol-3-OH kinase (PI(3)K), another effector of Ras⁶ that is important in both transformation²¹ and ES cell propagation^{22,23}. Co-immuno-

precipitation experiments showed that ERas, as well as HRasV12, associated with PI(3)K p110 δ (Fig. 4a). We also confirmed interaction between endogenous ERas and PI(3)K p85 (Supplementary Fig. S2c). Phosphorylation of Akt, a downstream molecule of PI(3)K, was decreased in *ERas*-null cells (Fig. 4b)^{24,25}. Transgenic re-expression of ERas partially recovered this reduction

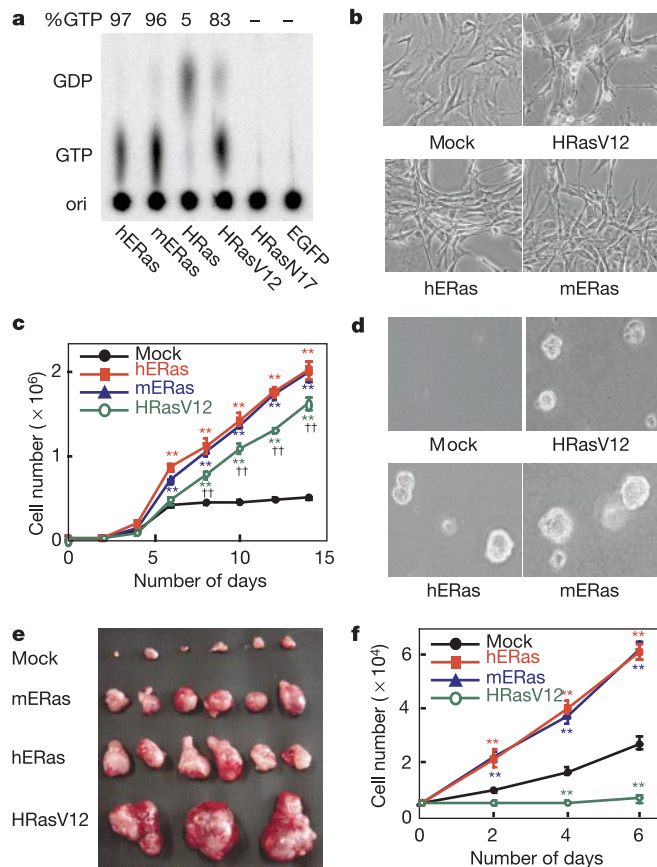


Figure 2 Transformation by ERas. **a**, Thin-layer chromatography showing protein association with GTP and GDP. **b**, Morphology of NIH 3T3 cells expressing HRasV12 or ERas. **c**, Cell growth. Ten thousand cells were plated and counted every other day for 14 d. Double asterisk, $P < 0.01$ versus mock transfected; double dagger, $P < 0.01$ versus ERas ($n = 4$). **d**, Anchorage-independent growth in soft agar. **e**, Tumours in nude mice. One million cells were injected subcutaneously and tumours were dissected after 14 d. **f**, Growth of MEFs expressing HRasV12 or ERas. Five thousand cells were plated and counted after 2, 4 and 6 d. Double asterisk, $P < 0.01$ versus mock transfected ($n = 5$).

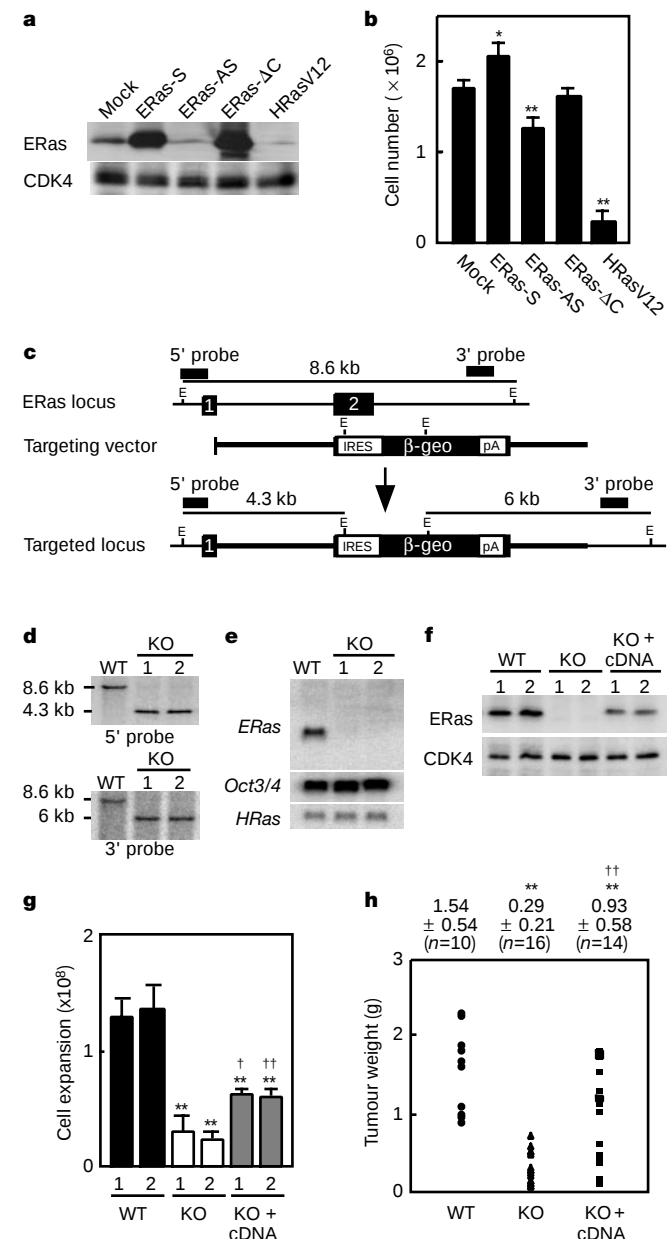


Figure 3 Role of ERas in ES cell growth. **a**, Western blot analysis of ERas expression in MG1.19 ES cells expressing sense cDNAs of ERas, ERas without the CAAX motif (ERas-ΔC) or HRasV12, or an antisense cDNA of ERas. **b**, Cell growth. Ten thousand of the cells in **a** were plated and counted after 6 d. Asterisk, $P < 0.05$ and double asterisk, $P < 0.01$ versus mock transfected ($n = 4$). **c**, Targeted disruption of *ERas*. Filled boxes indicate exons. The length of diagnostic *Eco*RI (E) restriction fragments and the locations of the 5' and 3' flanking probes are shown. **d**, Southern blot analyses. WT, wild-type ES cells; KO, *ERas*-knockout cells. **e**, Northern blot analyses. KO + cDNA, *ERas* knock-out cells expressing the *ERas* transgene. **f**, Western blot analysis. **g**, Cell expansion after 16 d of culture. Double asterisk, $P < 0.01$ versus wild type; dagger, $P < 0.05$ versus KO1; double dagger, $P < 0.01$ versus KO2 ($n = 4$). **h**, Teratoma formation. One million cells were subcutaneously injected into nude mice. Teratomas were dissected and weighed after 28 d. Double asterisk, $P < 0.01$ versus wild-type; double dagger, $P < 0.01$ versus KO.

(Supplementary Fig. S2d). These data show that ERas constitutively binds and activates PI(3)K.

We next examined whether impaired growth and tumorigenicity in *ERas*-null ES cells could be rescued by forced expression of active PI(3)K²⁶. Expression of the transgene was detected in two clones (one from each *ERas*-null clone) by northern blot analysis (data not shown). In both clones, the phosphorylation of Akt was higher than in wild-type cells (Fig. 4b) and impaired growth reverted to wild-type levels (Fig. 4c). Teratoma formation was also rescued and was even faster than in wild-type ES cells (Fig. 4d). In addition, we found that growth-promoting activity of ERas was blocked by the specific PI(3)K inhibitor LY294002, but not by the MAPK kinase inhibitor PD098059 (Supplementary Fig. S2e). These results show that the PI(3)K cascade is important in the growth-promoting activity of ERas.

Our study has established a previously unknown pathway that activates PI(3)K in ES cells. The role of PI(3)K in the cell-cycle control of ES cells has been shown by studies in which LY294002 markedly increased the proportion of ES cells in G0/G1 phase²⁷. By contrast, we observed only small changes in the cell cycle in *ERas*-null cells as compared with wild-type cells (G1, $24.6 \pm 0.3\%$; S, $60.5 \pm 1.5\%$; G2, $14.8 \pm 1.3\%$ versus G1, $22.8 \pm 0.7\%$; S,

$59.8 \pm 1.3\%$; G2, $17.4 \pm 0.8\%$), suggesting that PI(3)K might also have a growth-promoting effect that is independent of cell-cycle control. We have also shown that Akt is a downstream effector of the ERas/PI(3)K pathway, although other factors are also likely to be involved. The identification of ERas will facilitate our understanding of signalling pathways that maintain the tumour-like propagation of ES cells. □

Methods

Isolation of full-length cDNA of mouse *ERas*

We obtained the 5' portion of *ERas* cDNA using a 5' RACE system (version 2.0, Invitrogen) with primers 45328-AS1 for reverse transcription and 45328-race11 for polymerase chain reaction (PCR). All primer sequences are given in the Supplementary Information. Amplified products were cloned into pCR2.1 (Invitrogen) and four independent clones were sequenced.

Analysis of *HRas*p

A DNA fragment containing human *HRas*p was amplified from genomic DNA of two adults by PCR with the primers hHRAS2-S and hHRAS2-AS. The amplified product was cloned into pCR2.1 to construct pCR2.1-hERas and sequenced.

Analysis of GTP/GDP association

We transfected pCAG-IP-EGFP, pCAG-IP-Myc-hERas, pCAG-IP-Myc-mERas, pCAG-IP-Myc-HRas, pCAG-IP-Myc-HRasV12 or pCAG-IP-Myc-HRasN17 into MG1.19 cells by using Lipofectamine 2000 (Invitrogen). The construction of expression vectors is described in Supplementary Information. After 24 h, GTP/GDP association was evaluated as described²⁸, except that the lysis buffer consisted of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na₃VO₄, 20 mM MgCl₂, 0.5% Triton X-100 and protease inhibitor cocktail (Nacalai Tesque). Signals were analysed with a BAS 5000 imaging plate scanner (Fujifilm). The ratio of GTP-bound form was calculated as $GTP/(GTP + 1.5 \times GDP)$.

Retroviral infection

MEFs were isolated and cultured as described¹⁸. We cultured NIH 3T3 mouse fibroblast cells in Dulbecco's modified Eagle's medium supplemented with 10% calf serum. Cells of the ecotropic packaging line PLAT-E (8×10^6) were plated in a 100-mm dish, incubated for 24 h and then transfected using FuGENE6 (Roche) with a pMX-IP retroviral plasmid encoding ERas, ERas lacking the CAAX motif (ERas-ΔC) or HRasV12. After 24 h, the medium was replaced and incubation was continued for another 24 h. We then passed the virus-containing medium through a 0.45-μm filter (Schleicher & Schuell) and supplemented it with $4 \mu\text{g ml}^{-1}$ polybrene (Nacalai Tesque). We incubated MEFs (passage 1) or NIH 3T3 cells (~80% confluency) in this supernatant in a 100-mm dish for 24 h. The medium was then replaced and incubation was continued for another 24 h. Infected cell populations were selected with $2 \mu\text{g ml}^{-1}$ puromycin for 4 d. We did the transformation assay as described²⁹.

Episomal expression in MG1.19 ES cells

A sense strand cDNA of ERas, ERas-ΔC or HRasV12 or an antisense strand cDNA of ERas was inserted into a pCAG-IP vector containing the Polyoma replication origin. We transfected these plasmids with Lipofectamine 2000 (Invitrogen) into MG1.19 ES cells expressing Polyoma large antigen. Cells were selected with puromycin for 4 d.

Targeted disruption of mouse *ERas*

A targeting vector designed to replace the *ERas* coding region with β-geo was transferred into RF8 ES cells¹³ by electroporation. Targeted clones were identified by PCR and confirmed by Southern blot (Supplementary Information).

ERas and PI(3)K expression in *ERas*-deficient ES cells

pCAG-mERas-IH or pCAG-myr-p110-IH was transferred into *ERas*-deficient ES cells by electroporation. To identify clones expressing ERas, we screened colonies resistant to $100 \mu\text{g ml}^{-1}$ hygromycin B by western blot analysis with a rabbit antiserum against ERas, which was generated against histidine-tagged recombinant ERas produced in *Escherichia coli*. Clones expressing active PI(3)K were screened by northern blot analysis using PI(3)K p110α cDNA as a probe.

Measurement of expansion in cell numbers

We plated ES cells at 1×10^4 cells per well in 24-well plates. After 4 d, the cells were counted with a Z2 Coulter Counter (Beckman Coulter) and replated into new wells at the same density. This procedure was repeated for four passages. We calculated the expansion of cell numbers after a total of 16 d as follows: expansion in cell number = (cell number after 1st passage/ 1×10^4) × (cell number after 2nd passage/ 1×10^4) × (cell number after 3rd passage/ 1×10^4) × (cell number after 4th passage/ 1×10^4).

Immunoprecipitation

We transfected pCAG-IP-EGFP, pCAG-IP-Myc-hERas, pCAG-IP-Myc-mERas, pCAG-IP-Myc-HRas, pCAG-IP-Myc-HRasV12 or pCAG-IP-Myc-HRasN17 into MG1.19 cells, along with pCAG-IP-HA-Raf1, pCAG-IP-HA-BRaf or pCAG-IP-HA-PI(3)K-p110δ. Cell lysates were collected and immunoprecipitation were done as described³⁰. Precipitants and lysates were analysed by SDS-PAGE and western blot analysis. We used the following antibodies: agarose-conjugated monoclonal antibody against Myc (SC40

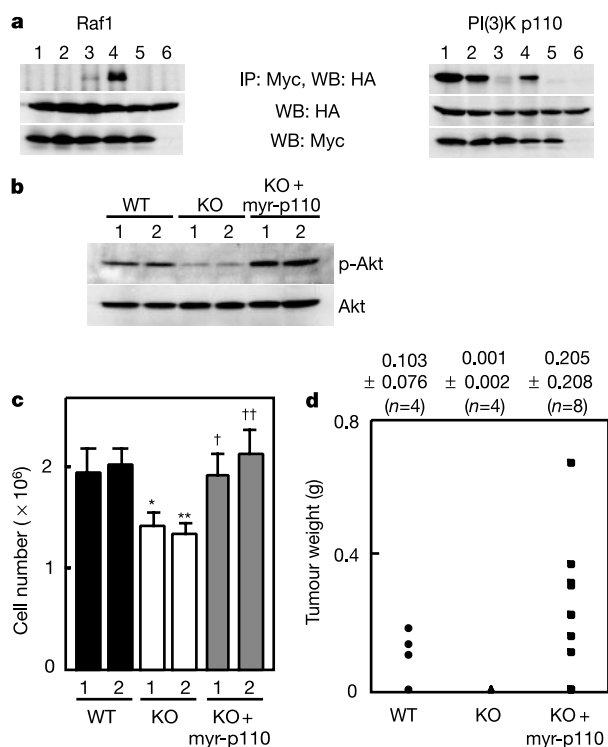


Figure 4 PI(3)K activation by ERas. **a**, Interaction with Raf1 and PI(3)K p110δ. MG1.19 cells were transfected with constructs expressing Myc-hERas (lane 1), Myc-mERas (lane 2), Myc-HRas (lane 3), Myc-HRasV12 (lane 4), Myc-HRasN17 (lane 5) or EGFP (lane 6), along with HA-Raf1 (left) or HA-PI(3)K (right). Cell lysates were analysed by western blot with antibodies against Myc (bottom) or HA (middle). Precipitants with agarose-conjugated anti-Myc antibodies were analysed with antibodies against HA (top). **b**, Akt phosphorylation. Wild-type ES cells (WT), *ERas*-null ES cells (KO) and *ERas*-null ES cells expressing active PI(3)K (KO + myr-p110) were analysed by western blot with antibodies against phosphorylated (Ser 473) or total Akt. **c**, Cell growth. Ten thousand of the cells in **b** were plated and counted after 6 d. Asterisk, $P < 0.05$ and double asterisk, $P < 0.01$ versus mock transfected; dagger, $P < 0.05$ versus KO1; double dagger, $P < 0.01$ versus KO2 ($n = 4$). **d**, Teratoma formation. One million of the cells in **b** were subcutaneously injected into nude mice. Teratomas were dissected and weighed after 16 d.

AC, Santa Cruz) for immunoprecipitation; a monoclonal antibody against haemagglutinin A (HA; 1867423, Roche), a polyclonal antibody against Myc (SC789, Santa Cruz), and antibodies against phosphorylated (Ser 473) or total Akt (9270, New England Biolabs).

Statistical analysis

Results shown are the mean $\pm$ s.d. We analysed data by one-way analysis of variance (ANOVA). Individual statistical differences were determined by Scheffé's multiple range comparison test.

Accession numbers

The sequences of mouse and human ERas can be retrieved from DDBJ/GenBank/EMBL with accession numbers AB093573 and AB093575.

Received 10 February; accepted 1 April 2003; doi:10.1038/nature01646.

1. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
2. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
3. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
4. Freed, C. R. Will embryonic stem cells be a useful source of dopamine neurons for transplant into patients with Parkinson's disease? *Proc. Natl Acad. Sci. USA* **99**, 1755–1757 (2002).
5. Seeburg, P. H., Colby, W. W., Capon, D. J., Goeddel, D. V. & Levinson, A. D. Biological properties of human *c-Ha-ras* genes mutated at codon 12. *Nature* **312**, 71–75 (1984).
6. Rodriguez-Viciana, P. *et al.* Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature* **370**, 527–532 (1994).
7. Moodie, S. A., Willumsen, B. M., Weber, M. J. & Wolfman, A. Complexes of Ras.GTP with Raf-1 and mitogen-activated protein kinase kinase. *Science* **260**, 1658–1661 (1993).
8. Zhang, X. F. *et al.* Normal and oncogenic p21^{ras} proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364**, 308–313 (1993).
9. Takai, Y., Sasaki, T. & Matozaki, T. Small GTP-binding proteins. *Physiol. Rev.* **81**, 153–208 (2001).
10. Chen, Z. Q., Ulsh, L. S., DuBois, G. & Shih, T. Y. Posttranslational processing of p21 ras proteins involves palmitoylation of the C-terminal tetrapeptide containing cysteine-186. *J. Virol.* **56**, 607–612 (1985).
11. Willumsen, B. M., Christensen, A., Hubbard, N. L., Papageorge, A. G. & Lowy, D. R. The p21 ras C-terminus is required for transformation and membrane association. *Nature* **310**, 583–586 (1984).
12. Miyoshi, J., Kagimoto, M., Soeda, E. & Sakaki, Y. The human *c-Ha-ras2* is a processed pseudogene inactivated by numerous base substitutions. *Nucleic Acids Res.* **12**, 1821–1828 (1984).
13. Meiner, V. L. *et al.* Disruption of the acyl-CoA:cholesterol acyltransferase gene in mice: evidence suggesting multiple cholesterol esterification enzymes in mammals. *Proc. Natl Acad. Sci. USA* **93**, 14041–14046 (1996).
14. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
15. Nichols, J., Evans, E. P. & Smith, A. G. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341–1348 (1990).
16. Gassmann, M., Donoho, G. & Berg, P. Maintenance of an extrachromosomal plasmid vector in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **92**, 1292–1296 (1995).
17. Fasano, O. *et al.* Analysis of the transforming potential of the human H-ras gene by random mutagenesis. *Proc. Natl Acad. Sci. USA* **81**, 4008–4012 (1984).
18. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
19. Cheng, A. M. *et al.* Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* **95**, 793–803 (1998).
20. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30–43 (1999).
21. Rodriguez-Viciana, P. *et al.* Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. *Cell* **89**, 457–467 (1997).
22. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. Pten is essential for embryonic development and tumour suppression. *Nature Genet.* **19**, 348–355 (1998).
23. Sun, H. *et al.* PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5-trisphosphate and Akt/protein kinase B signaling pathway. *Proc. Natl Acad. Sci. USA* **96**, 6199–6204 (1999).
24. Burgering, B. M. & Coffer, P. J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
25. Franke, T. F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727–736 (1995).
26. Klippel, A. *et al.* Membrane localization of phosphatidylinositol 3-kinase is sufficient to activate multiple signal-transducing kinase pathways. *Mol. Cell. Biol.* **16**, 4117–4127 (1996).
27. Jirmanova, L., Afanassieff, M., Gobert-Gosse, S., Markossian, S. & Savatier, P. Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells. *Oncogene* **21**, 5515–5528 (2002).
28. Quilliam, L. A. M.-R. *et al.* Ras/R-Ras3, a transforming ras protein regulated by Sos1, GRF1, and p120 Ras GTPase-activating protein, interacts with the putative Ras effector AF6. *J. Biol. Chem.* **274**, 23850–23857 (1999).
29. Clark, G. J., Cox, A. D., Graham, S. M. & Der, C. J. Biological assays for Ras transformation. *Methods Enzymol.* **255**, 395–412 (1995).
30. Rosario, M., Paterson, H. F. & Marshall, C. J. Activation of the Raf/MAP kinase cascade by the Ras-related protein TC21 is required for the TC21-mediated transformation of NIH 3T3 cells. *EMBO J.* **18**, 1270–1279 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Kaiho, Y. Tokuzawa and M. Murakami for discussion; C. Takigawa and J. Iida for technical assistance; T. Ichisaka and Y. Samitsu for blastocyst microinjection; R. Farese Jr for RF8 ES cells, R. Jaenisch and T. Noda for J1 cells; W. Skarnes and S. Young for CGR8 cells; H. Niwa for MG1.19 cells, pCAG-IP and pBIKS(–)BgeopA; M. Okabe and J.-i. Miyazaki for pCX–EGFP; K. Kohno and T. Kitamura for PLAT-E cells and pMX retroviral vectors; and S. Young, R. Farese Jr and R. Pitas for critically reading the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Y. (shinyay@gtc.aist-nara.ac.jp).

Modulation of oestrogen receptor signalling by association with the activated dioxin receptor

Fumiaki Ohtake*, Ken-ichi Takeyama*†, Takahiro Matsumoto*, Hirochika Kitagawa*, Yasuji Yamamoto‡, Keiko Nohara§||, Chiharu Tohyama§||, Andree Krust¶, Junsei Mimura||#, Pierre Chambon¶, Junn Yanagisawa*†, Yoshiaki Fujii-Kuriyama||# & Shigeaki Kato*†

* The Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-0032, Japan
† SORST, Japan Science and Technology, Kawaguchi, Saitama 332-0012, Japan
‡ Taiho Pharmaceutical Company Ltd, Cancer Research Laboratory, Hanno Research Center, Hanno, Saitama, 357-8527, Japan
§ National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan
|| CREST, Japan Science and Technology, Kawaguchi, Saitama 332-0012, Japan
¶ Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, Université Louis Pasteur, Collège de France, 67404 Illkirch, Strasbourg, France
TARA Center, University of Tsukuba, Tennodai, Tsukuba, 305-8577, Japan

Environmental contaminants affect a wide variety of biological events in many species. Dioxins are typical environmental contaminants that exert adverse oestrogen-related effects¹. Although their anti-oestrogenic actions^{2,3} are well described, dioxins can also induce endometriosis^{4–7} and oestrogen-dependent tumours^{8,9}, implying possible oestrogenic effects. However, the molecular mechanism underlying oestrogen-related actions of dioxins remains largely unknown. A heterodimer of the dioxin receptor (AhR) and Arnt, which are basic helix–loop–helix/PAS-family transcription factors, mediates most of the toxic effects of dioxins^{10,11}. Here we show that the agonist-activated AhR/Arnt heterodimer directly associates with oestrogen receptors ER- α and ER- β . This association results in the recruitment of unliganded ER and the co-activator p300 to oestrogen-responsive gene promoters, leading to activation of transcription and oestrogenic effects. The function of liganded ER is attenuated. Oestrogenic actions of AhR agonists were detected in wild-type ovariectomized mouse uteri, but were absent in AhR^{–/–} or ER- α ^{–/–} ovariectomized mice. Our findings suggest a novel mechanism by which ER-mediated oestrogen signalling is modulated by a co-regulatory-like function of activated AhR/Arnt, giving rise to adverse oestrogen-related actions of dioxin-type environmental contaminants.

ERs, which are members of the nuclear receptor (NR) family^{12,13}, and AhR/Arnt are both ligand-dependent transcription factors. Ligand-activated AhR heterodimerizes with Arnt and activates the transcription of dioxin target genes such as *CYP1A1* (refs 10,11) through xenobiotic response elements (XREs). ERs bind to oestrogen response elements (EREs) and activate transcription in an oestrogen-dependent manner. This transcriptional activation

requires the recruitment of co-activator complexes^{13–18}, including histone acetyltransferase (HAT) complexes containing p300 and CREB binding protein (CBP). In view of previous reports that AhR ligands exhibit oestrogen-related adverse effects, it is possible that ER-mediated oestrogen signalling might cross-talk with AhR-mediated signalling through an unknown mechanism that regulates transcription. We therefore decided to examine whether AhR/Arnt heterodimer could transcriptionally affect ER transactivation functions, thereby modulating oestrogen signalling.

To monitor the transactivation function of endogenous receptors, luciferase reporter plasmids bearing consensus binding elements—ERE for ERs, and XRE for AhR/Arnt—were transfected into MCF-7 cells, a breast cancer cell line known to express both receptors endogenously². Although the synthetic AhR ligand 3-methylcholanthrene (3MC) effectively activated transcription through XRE¹⁰, 17 β -estradiol (E2) did not, as expected (Fig. 1a). However, to our surprise, 3MC alone activated ERE-mediated transcription in the absence of E2 (Fig. 1a). In the presence of E2, ERE-mediated transcription was decreased by the addition of 3MC. Western blotting showed that the amount of ligand-induced transactivation did not simply reflect variations in receptor numbers (Fig. 1a). 3MC alone decreased AhR and ER- α protein levels, in agreement with previous reports¹⁹.

We then examined the effect of AhR/Arnt on ER-mediated transcription by using exogenous receptors in Ishikawa cells, a uterine tumour cell line. Again, 3MC potently stimulated ERE-mediated transcription in the absence of E2 when both ER (either ER- α or ER- β) and AhR/Arnt were expressed, whereas it lowered the

E2-induced transactivation function of ERs (Fig. 1b) without binding directly to ERs (Supplementary Fig. 1a) or affecting expression levels of ERs (data not shown). This activation effect of 3MC requires ERE (Fig. 1b, lanes 1–4), ER- α (lanes 7 and 8), AhR (lanes 9 and 10) and Arnt (lanes 11 and 12). To verify that an AhR ligand does indeed exert oestrogenic action through direct binding to AhR, other AhR ligands were further tested. More stable ligands such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), benzo[*a*]pyrene and β -naphthoflavone acted as agonists, like 3MC (Fig. 1c), whereas the oestrogenic action of 3MC was blocked by either a known AhR antagonist, α -naphthoflavone or a pure oestrogen antagonist, ICI182,780 (Fig. 1d). The modulation of transcription activity by AhR/Arnt observed with ERs was not detected on other NRs including glucocorticoid receptor, progesterone receptor, vitamin D receptor (VDR), retinoic acid receptor and peroxisome proliferator activated receptor- γ (PPAR- γ) (data not shown).

Because ERs possess two transactivation functions, AF-1 and AF-2, in the amino-terminal A/B and carboxy-terminal E/F regions, respectively^{16,20}, we examined the functional association of AhR/Arnt with these two regions using ER deletion mutants (HE15 for AF-1 domain, and HE19 for AF-2 domain) (Supplementary Fig. 1b) in Ishikawa cells. The N-terminal A/B regions of ER- α and ER- β were required for stimulation of ERE-mediated transcription by AhR/Arnt, whereas we detected no modulation of AF-2 functions (Fig. 1e)²⁰. Thus, 3MC-bound AhR/Arnt might modulate the functions of ERs through association with the N-terminal A/B regions. This possibility was supported by the observation that

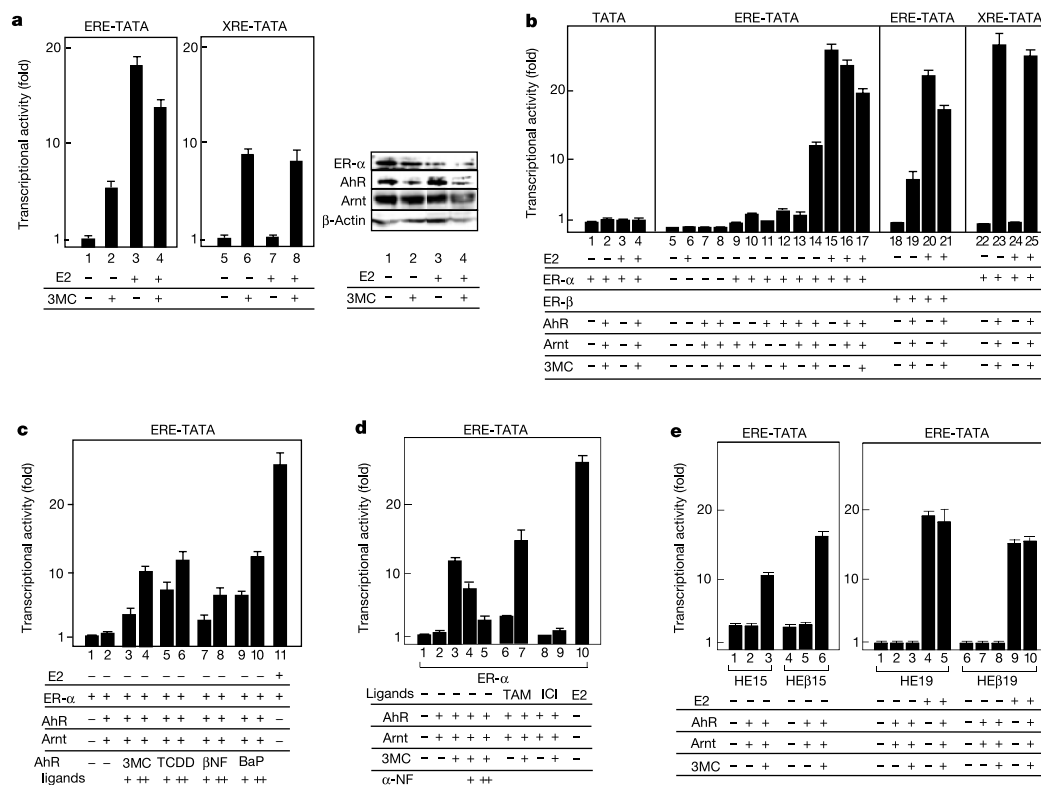


Figure 1 Activation of unliganded ER function by liganded dioxin receptor heterodimer. **a**, A dioxin receptor ligand activates transcription mediated through an ERE. MCF-7 cells were transfected with the reporter plasmids ERE-luciferase or XRE-luciferase in the presence or absence of E2 (10 nM) and 3MC (1 μ M). Luciferase assays were performed with the cell extracts. All values are means $\pm$ s.d. for at least three independent experiments. **b**, Liganded AhR/Arnt induces the transactivation function of ERE-bound unliganded ER. Ishikawa cells transfected with the indicated plasmids were subjected to

luciferase assays. **c**, Transactivation of unliganded ER by the other AhR agonists. **d**, Potentiation of ERE-mediated transcription by liganded AhR/Arnt is blocked by an antagonist for either ER- α or AhR. Cells treated with tamoxifen (TAM; 100 nM), ICI182,780 (ICI; 100 nM), 3MC (+, 100 nM; ++, 1 μ M), TCDD (+, 10 nM; ++, 100 nM), β -naphthoflavone (β -NF; +, 100 nM; ++, 1 μ M), benzo[*a*]pyrene (BaP; +, 10 nM; ++, 100 nM), α -naphthoflavone (α -NF; +, 100 nM; ++, 1 μ M). **e**, Potentiation of ERE-mediated transcription by AhR/Arnt is mediated by the ERs A/B regions.

3MC-bound AhR/Arnt potentiates the transactivation function of ER- α in the presence of the ER- α AF-1 agonist/AF-2 antagonist tamoxifen (Fig. 1d)¹⁶.

We then tested whether a 3MC-dependent physical interaction occurred between AhR/Arnt and ERs. Irrespective of E2 binding, endogenous ER- α in MCF-7 cells, and tagged ER- α overexpressed in COS-1 cells, were found to co-immunoprecipitate with 3MC-bound AhR, but not with unliganded AhR, only when Arnt was co-expressed (Fig. 2a and b). In agreement with the functional interaction between AhR/Arnt and the A/B region of ER- α (Fig. 1e), a 3MC-dependent interaction between AhR/Arnt and HE15 was observed, but not between AhR/Arnt and HE19 (ref. 12). Although ER- β , like ER- α , also associated with AhR in a 3MC-dependent fashion, no other receptors tested showed such an association (Fig. 2b).

Moreover, a direct interaction between AhR, but not Arnt, and A/B regions from both ER- α and ER- β could be mapped by an *in vitro* glutathione S-transferase (GST) pull-down assay (Fig. 2c). It therefore seems that, upon ligand binding and nuclear translocation¹⁰, AhR heterodimerizes with nuclear Arnt and then associates with unliganded ER- α or ER- β , which are constitutively in the nucleus¹⁶, through direct interaction with their A/B regions. Further

analyses by GST pull-down assay mapped the small regions of the A/B region of ER- α (residues 40–120), the A/B region of ER- β (residues 33–55)²¹, and the helix–loop–helix/PAS domain of AhR²², which are indispensable for direct interaction *in vitro* (Fig. 2d and Supplementary Fig. 2a). An ER- α mutant lacking the AhR-interacting region (ER- α Δ AhR) failed to be activated by AhR/Arnt but responsiveness to E2 was still retained, supporting the idea that the interaction is required for AhR ligand-induced activation of the ER function (Fig. 2e).

To explore the molecular mechanisms of the 3MC-dependent transactivation function of AhR and ERs, we used co-immunoprecipitation to examine whether p300 was recruited to the complex, because both AhR and ERs have been independently reported to require p300/CRB as a co-activator^{10,16,18,23}. p300 was recruited to ER- α in the presence but not the absence of E2 (Fig. 2f, lanes 2 and 4). However, even in the absence of E2, p300 associated with 3MC-bound AhR/Arnt and unliganded ER- α to form a complex (Fig. 2f, lane 3). Recruitment of the p160 family co-activator SRC-1 (ref. 13; Fig. 2f, lane 3), TIF2 or AIB1 (data not shown) to AhR/Arnt-associated ERs were not detected. Thus, the co-activator complex required to activate transcription by the unliganded ERs associated with liganded AhR/Arnt might be distinct from both co-activator

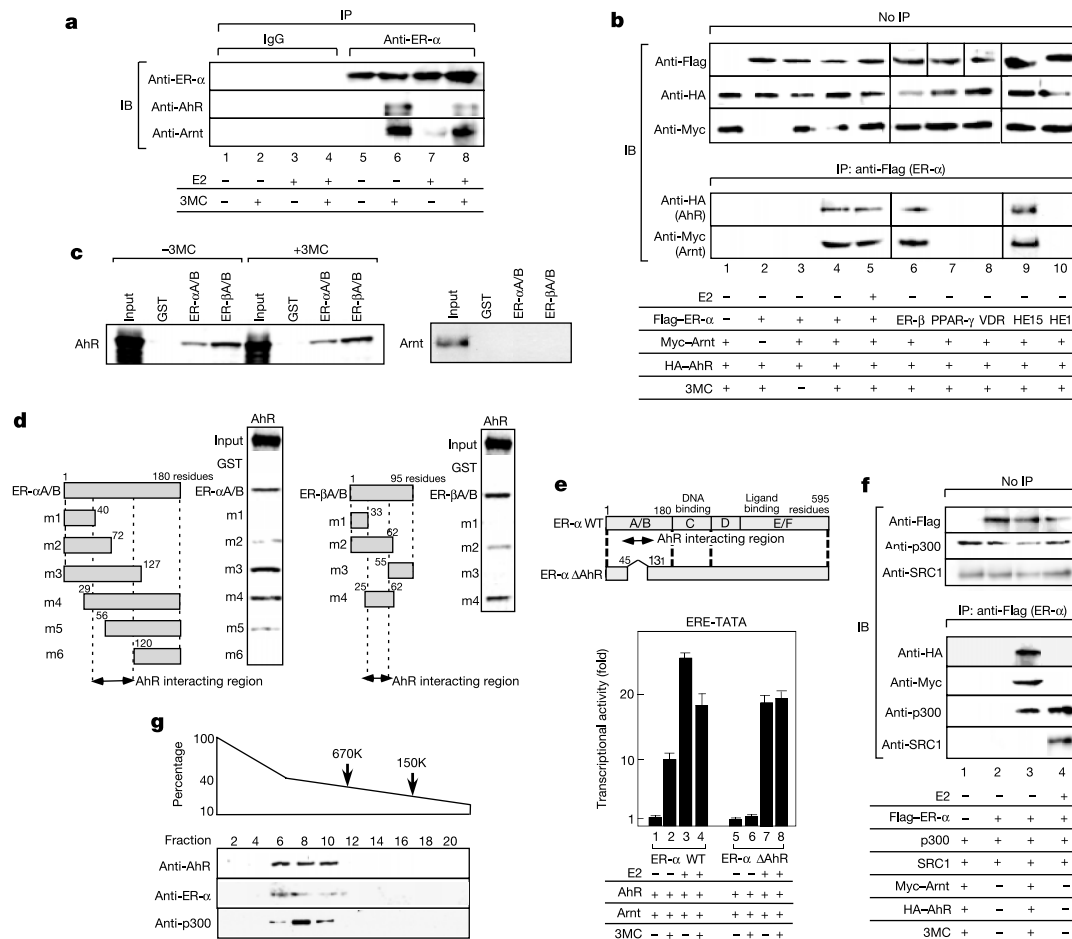


Figure 2 3MC-dependent interaction of ERs with AhR/Arnt. **a**, 3MC-dependent but E2-independent interaction of endogenous ER- α with AhR/Arnt in MCF-7 cells. Cells were subjected to immunoprecipitation (IP) with mouse anti-ER- α or normal mouse immunoglobulin as a control. The immunoprecipitates were western blotted (IB) with specific antibodies as indicated. **b**, E2-independent, 3MC-dependent interaction of exogenous ERs with AhR/Arnt in COS-1 cells. The transfected cells were subjected to immunoprecipitation and then western blotting. PPAR, peroxisome proliferator-activated receptor; VDR, vitamin D receptor. **c**, Direct but 3MC-independent interaction of AhR with

ER- α and ER- β in an *in vitro* GST pull-down assay. **d**, Mapping the interaction domains of ER- α and ER- β with AhR. **e**, The AhR-interacting core region in the ER- α A/B domain is required for ER- α activation by AhR/Arnt. Luciferase assays with the indicated ER derivative. **f**, Recruitment of p300 co-activator to a complex containing unliganded ER- α and 3MC-bound AhR/Arnt. **g**, AhR/ER- α /p300 form a complex on glycerol gradient analysis. The Flag-AhR associated proteins in stable transformant HeLa cells were fractionated by molecular mass by a glycerol gradient assay.

complexes for the unassociated receptors. Indeed, ER- α and p300 were detected in the same fractions as Flag ([EYKEEEK]₂)-tagged AhR fractionated by a glycerol gradient, suggesting that they form a complex with a relative molecular mass (M_r) larger than 670,000 (670K) (Fig. 2g).

To investigate whether the observed association between AhR and ERs occurred on EREs in endogenous target gene promoters of MCF-7 cells, we performed a chromatin immunoprecipitation (ChIP) analysis with pS2 and *c-fos* gene promoters¹⁷. Interestingly,

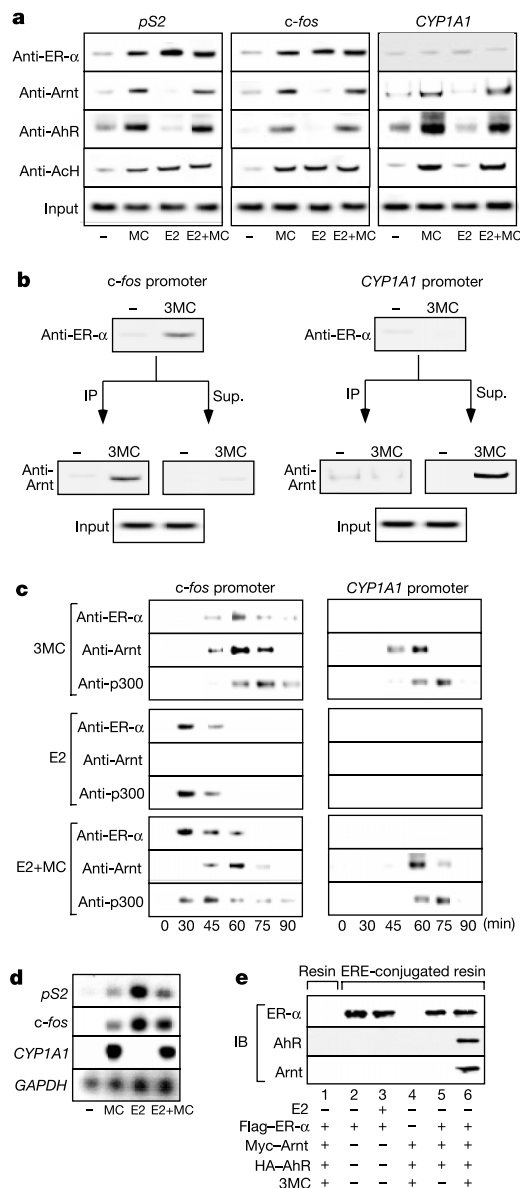


Figure 3 3MC-dependent recruitment of AhR/Arnt to ER- α bound on oestrogen-responsive gene promoters. **a**, 3MC-dependent interaction with AhR/Arnt induces ERE binding of unliganded ER- α to E2 responsive gene promoters in MCF-7 cells. For ChIP analyses, soluble chromatin prepared from MCF-7 cells treated with ligands for 45 min was immunoprecipitated with the indicated antibodies. The final DNA extracts were amplified using specific sets of primer pairs to detect the *c-fos*, pS2 and *CYP1A1* gene promoters as indicated. **b**, 3MC-dependent association of AhR/Arnt with ER- α bound to E2-responsive gene promoters. The immunoprecipitates and their supernatants were sequentially applied for ChIP analysis as indicated. **c**, Dynamics of ER- α -Arnt-p300 assembly on ligand-responsive gene promoters. Occupancy of the *c-fos* and *CYP1A1* promoters by ER- α , Arnt and p300 at different times after ligand treatments. **d**, Induction of target genes examined by northern blot analysis. **e**, Complex formation of AhR-Arnt-ER- α on ERE through ER- α as revealed by ABC assay.

3MC induced binding of ER- α to ERE, as did E2, with AhR/Arnt recruitment. As expected, 3MC induced the recruitment of AhR/Arnt, but not ER- α , to the *CYP1A1* promoter XRE (Fig. 3a). Reflecting the recruitment of the receptors, acetylation of histone H4 was observed in the promoters (Fig. 3a), indicating the possible recruitment of a HAT co-activator complex to the receptors. The expression of these genes was accordingly induced by 3MC or E2 (Fig. 3d). Thus, the 3MC-dependent association between AhR/Arnt and ER- α seems to promote the binding of unliganded ER- α to EREs.

A ChIP assay involving sequential immunoprecipitation confirmed the 3MC-dependent association of AhR/Arnt with ER- α on

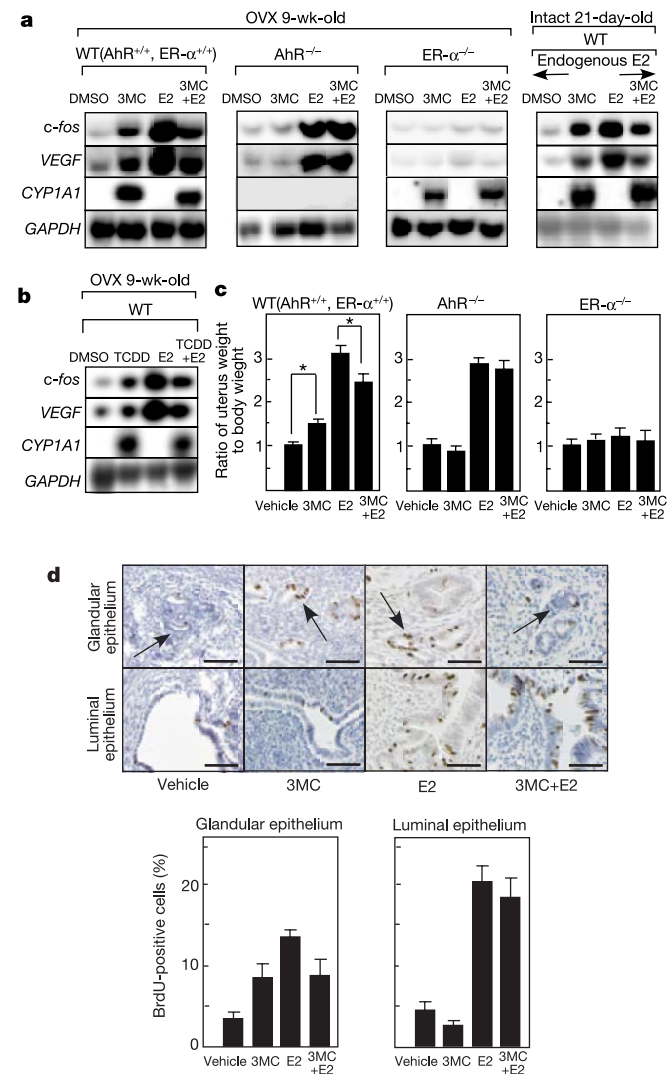


Figure 4 Oestrogenic actions of 3MC in mouse uterus are mediated by AhR and ER- α . **a**, Induction of E2-responsive genes by AhR agonists is mediated by both AhR and ER- α . Nine-week-old ovariectomized (OVX) mice and intact 21-day-old female mice of the indicated genotypes were injected with the ligands. Three hours later, total RNA was extracted from the uterus, then subjected to northern blot analysis with cDNAs for the target genes for E2 (*c-fos*, *VEGF*) and for 3MC (*CYP1A1*); *GAPDH* cDNA was used as an internal control. WT, wild type. **b**, The 3MC-induced increase in uterine wet weight (measured as the ratio of uterine wet weight in milligrams to body weight in grams) in ovariectomized mice was abolished by inactivation by either AhR or ER- α . The *t*-test shows a significant difference ($P < 0.01$) between 3MC-treated ($n = 9$) and olive-oil-treated ($n = 9$) wild-type mice. There is no significant difference ($P > 0.2$) between 3MC-treated ($n = 4$) and olive-oil-treated ($n = 4$) animals in either AhR^{-/-} and ER- α ^{-/-} mice. All values are means $\pm$ s.e.m. **d**, Induction of endometrial cell proliferation by 3MC and E2. BrdU-positive cells (brown) are indicated by arrows.

ERE (Fig. 3b). A time-course ChIP assay showed that ER- α , AhR and p300 HAT were simultaneously recruited to the *c-fos* promoter, presumably upon the binding of 3MC to AhR (Fig. 3c).

To verify the interaction of AhR/Arnt with ER- α bound to ERE in the promoters, the formation of a complex with ERE was tested by avidin-biotin-conjugated DNA(ERE) (ABCD) precipitation²⁴ (Fig. 3e). ER- α bound to consensus ERE (Fig. 3e, lanes 2, 3, 5 and 6), whereas AhR/Arnt alone did not (Fig. 3e, lane 4). However, in the presence of ER- α , AhR/Arnt was recruited to ERE in a 3MC-dependent manner (Fig. 3e, lanes 5 and 6). In the transient luciferase assay, the binding of ER- α to ERE and the activation function of both AhR and Arnt were required for the activation of ER- α through ERE by AhR/Arnt (Supplementary Fig. 3a, lanes 3, 7, and 8), whereas the AF-1 and AF-2 activities of ER- α and the DNA-binding capacity of the AhR/Arnt heterodimer were dispensable (Supplementary Fig. 3a, lanes 4–6).

Finally, we tested whether AhR-ligand-dependent AhR-ER interaction was responsible for the oestrogenic actions of AhR agonists in the absence of oestrogens on gene expression in intact animals. In addition to the induction of the *CYP1A1* gene, treatments with 3MC (Fig. 4a) and TCDD (Fig. 4b) for 3 hours stimulated the expression of the oestrogen-responsive genes *c-fos*²⁵ and *vascular endothelial growth factor* (*VEGF*)²⁶ in the uteri of ovariectomized wild-type mice (Fig. 4a, b). This oestrogenic action of 3MC in the uterus was also detected in intact 21-day-old female mice, whereas the AhR agonists exhibited anti-oestrogenic activities in the presence of high doses of oestrogen (Fig. 4a). There have been conflicting reports on the induction of *c-fos* by AhR ligands: one is that AhR ligands repress the E2-induced expression of *c-fos*³; the other is that AhR ligands themselves induce the expression of *c-fos*²⁷. The 3MC-mediated activation of oestrogen target genes was completely abolished in both AhR^{-/-} (ref. 28) and ER- α ^{-/-} (ref. 29) ovariectomized mice, although each receptor knockout mouse strain retained ligand responsiveness (Fig. 4a) and the expression (Supplementary Fig. 4a) of the other intact receptor. The injection of 3MC led to increases in uterine wet weight, as did that of E2 (Fig. 4c). This action of 3MC was again abolished in both AhR^{-/-} and ER- α ^{-/-} mice (Fig. 4c).

To examine whether the increased uterine wet weight was due to the proliferation of endometrial cells, DNA synthesis in uterine epithelial cells was examined by labelling with bromodeoxyuridine (BrdU). Ovariectomized mice treated with 3MC exhibited enhanced cell proliferation in the glandular epithelium, as did E2-treated mice (Fig. 4d). Proliferation of the luminal epithelium was enhanced by E2 but not by 3MC.

The present findings indicate that the oestrogenic action of AhR agonists might be exerted through a direct interaction between AhR/Arnt and unliganded ER and by the formation of functional units bound to EREs that activate transcription, at least in uterine gene induction and cellular proliferation. The most marked manifestation of the possible oestrogenicity of dioxins could be seen as their linking to endometriosis^{4–7}, because oestrogen is the major factor in the stimulation of proliferation of these cells. Thus, AhR expressed in the uterine glandular epithelium³⁰ might respond to dioxins by associating with unliganded ERs, which then stimulates oestrogen-dependent cell proliferation. In contrast, AhR agonists exhibit anti-oestrogenic activities in the presence of high doses of E2 in animals³ and cultured cell lines². We also found that AhR/Arnt repressed E2-bound ER function, which is consistent with these previous reports. However, whereas most previous studies have not examined or mentioned the effects of AhR ligands in the absence of E2, we addressed this issue carefully in the present study. Thus, oestrogen concentrations, which vary with age, oestrous cycle, tissues and other factors, might define the oestrogenic/anti-oestrogenic actions of the AhR ligands in intact animals. Our present model, in which AhR potentiates unliganded ERs but represses liganded ER, might be an explanation of these previous findings,

and it will be of interest to identify the other components of the liganded AhR-ER- α complex involved in the oestrogenic/anti-oestrogenic actions of dioxins. Our proposal is that one of the molecular mechanisms for the oestrogen-related adverse effects of dioxin-type environmental contaminants is the modulation of oestrogen receptor signalling by dioxin-dependent association with dioxin receptor. □

Methods

Plasmids

Full-length complementary DNAs of AhR and Arnt were inserted into pcDNA3 vectors (Invitrogen). Three consensus EREs¹⁶ and XREs²² were inserted into the promoter of luciferase pGL3-basic vector to generate ERE-TATA-luciferase and XRE-TATA-luciferase, respectively. ER- α Δ AhR was generated by the deletion of 45–131 residues from ER- α . The other mutants of ER- α and ER- β were as described previously²¹.

Transfection and luciferase assay

Human endometrium cancer-derived Ishikawa cells, human breast cancer-derived MCF-7 cells, green monkey COS-1 cells and human 293T cells maintained in DMEM supplemented with 10% FBS were cultured in phenol-red-free DMEM containing 0.2% charcoal-stripped FBS before assays. Cells at 40–50% confluence were transfected with the indicated plasmids (0.25 μ g ERE-Luc, 0.1 μ g XRE-Luc, 0.025 μ g ER- α , AhR and Arnt were transfected) using Lipofectamine reagent (Gibco BRL) in 12-well Petri dishes. Total amounts of cDNA were adjusted by supplementing with empty vector up to 1.0 μ g. Cells were treated with E2 (100 nM) and 3MC (1 μ M). Luciferase activity was determined with the Luciferase Assay System (Promega)¹⁶. As a reference plasmid to normalize transfection efficiency, 25 ng pRL-CMV plasmid (Promega) was co-transfected in all experiments. Results are given as means $\pm$ s.d. for at least three independent experiments.

Immunoprecipitation and GST pull-down assay

Whole cell extracts¹⁷ were used for immunoprecipitation with either anti-ER- α or anti-Flag antibody (anti-ER- α Ab-4 from Neo Markers; anti-Flag from Santa Cruz Biotechnology) after western blotting with anti-ER- α (Chemicon), anti-Arnt (Santa Cruz Biotechnology), anti-AhR (Santa Cruz Biotechnology), anti-p300 (Upstate Biotechnology), anti-SRC-1 (Santa Cruz Biotechnology), anti-Flag, anti-haemagglutinin and anti-Myc (Invitrogen) antibodies. Normal mouse immunoglobulin was used as a control. For immunoprecipitation of overexpressed proteins, cells were transfected as indicated with Flag-tagged ERs (5 μ g), haemagglutinin-tagged AhR (3 μ g), Myc-Arnt (5 μ g), SRC-1 (0.7 μ g) and p300 (0.7 μ g) in the presence or absence of 3MC and E2. For the GST pull-down assay, AhR and Arnt were translated *in vitro* and incubated with either GST, GST-ER- α (A/B) or GST-ER- β (A/B) immobilized on glutathione-Sepharose beads¹⁷.

Purification and separation of AhR-interacting complexes

HeLa nuclear extracts were loaded on an M2 anti-Flag agarose gel (Kodak). After being washed with binding buffer, the bound proteins were eluted from the agarose by incubation overnight with 2.5–5.0 ml of the Flag peptide (Kodak) in binding buffer (0.2 mg ml⁻¹). For fractionation on a glycerol gradient, eluents were layered on the top of a 13-ml linear 100–10% glycerol gradient and centrifuged for 16 h at 40,000 r.p.m. in an SW40 rotor (Beckman). Each fraction was western blotted with anti-AhR, anti-ER- α and anti-p300 antibodies. The protein standards used were β -globulin (M_r 158K) and thyroglobulin (667K)¹⁷.

Chromatin immunoprecipitation

Soluble chromatins of MCF-7 cells prepared with the acetyl-histone H4 immunoprecipitation assay kit (Upstate Biotechnology) were immunoprecipitated with antibodies against the indicated proteins. Specific primer pairs were designated to amplify the promoter regions of the *c-fos* (5'-GAAGAGTGGAGAAGGG-3' and 5'-GAAGCTGTGCTTACGG-3'), *pS2* (5'-AAAGAATTAGCTTAGGCC-3' and 5'-ACCTTAATCCAGGTCC-3') or *CYP1A1* (5'-CTTCGCCATCCATTCC-3' and 5'-GGGACTCCTCTTCGAC-3') genes from the extracted DNA. Optimal PCR conditions to allow semiquantitative measurement were used on 2% agarose/Tris-acetate-EDTA gels¹⁷. As a usual condition, cells were treated with ligands for 45 min. The inductions of the target genes were examined by northern blot analysis in MCF-7 cells treated with the ligands for 3 h.

ABCD precipitation

Avidin resin (Promega) was incubated with biotin-conjugated consensus ERE oligonucleotides, followed by incubation with cell lysates in lysis buffer (20 mM HEPES, 100 mM KCl, 0.5 mM EDTA, 0.1% Triton X-100 and 1 mM dithiothreitol) for 30 min. The subsequent ERE-protein complexes trapped on the resin were then eluted and western blotted²⁴.

Oestrogen responses in uterus

Nine-week-old female C57BL/6 mice with the indicated genotypes were ovariectomized. After 2 weeks the mice were treated with 3MC (4 mg kg⁻¹), TCDD (40 μ g kg⁻¹), and/or E2 (20 μ g kg⁻¹) in olive oil for 3 h. Total RNA was extracted from the uteri by Isogen (Wako Co.) and then subjected to northern blot analysis with cDNAs for the target genes for E2 (*c-fos*, *VEGF*) and for 3MC (*CYP1A1*), with *GAPDH* cDNA (encoding glyceraldehydes-3-

phosphate dehydrogenase) as an internal control¹⁶. For experiments with intact mice, 21-day-old female mice were used.

For uterine weight analysis, mice were treated with ligands for 3 days, and the ratio of uterine wet weight to body weight was calculated, followed by *t*-test analysis. Results are given as means $\pm$ s.e.m.

For the BrdU labelling experiment, ovariectomized mice were treated with ligands for 3 days, then injected with BrdU (30 mg kg⁻¹). Paraffin sections from the uteri 8 h after BrdU injection were immunostained with anti-BrdU monoclonal antibody by using the BrdU Labeling and Detection Kit 1 (Roche), and the percentage of BrdU-positive epithelial cells in the sections was calculated.

Received 19 February; accepted 1 April 2003; doi:10.1038/nature01606.

1. Bock, K. W. Aryl hydrocarbon or dioxin receptor: biologic and toxic responses. *Rev. Physiol. Biochem. Pharmacol.* **125**, 1–42 (1994).
2. Krishnan, V. *et al.* Molecular mechanism of inhibition of estrogen-induced cathepsin D gene expression by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in MCF-7 cells. *Mol. Cell. Biol.* **15**, 6710–6719 (1995).
3. Astroff, B. *et al.* Inhibition of the 17 β -estradiol-induced and constitutive expression of the cellular protooncogene *c-fos* by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in the female rat uterus. *Toxicol. Lett.* **56**, 305–315 (1991).
4. Gibbons, A. Dioxin tied to endometriosis. *Science* **262**, 1373 (1993).
5. Mayani, A., Barel, S., Soback, S. & Almogor, M. Dioxin concentrations in women with endometriosis. *Hum. Reprod.* **12**, 373–375 (1997).
6. Rier, S. E., Martin, D. C., Bowman, R. E., Dmowski, W. P. & Becker, J. L. Endometriosis in rhesus monkeys (*Macaca mulatta*) following chronic exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. *Fundam. Appl. Toxicol.* **21**, 433–441 (1993).
7. Cummings, A. M., Metcalf, J. L. & Birnbaum, L. Promotion of endometriosis by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in rats and mice: time-dose dependence and species comparison. *Toxicol. Appl. Pharmacol.* **138**, 131–139 (1996).
8. Brown, N. M., Manzolillo, P. A., Zhang, J. X., Wang, J. & Lamartiniere, C. A. Prenatal TCDD and predisposition to mammary cancer in the rat. *Carcinogenesis* **19**, 1623–1629 (1998).
9. Davis, B. J., McCurdy, E. A., Miller, B. D., Lucier, G. W. & Tritscher, A. M. Ovarian tumors in rats induced by chronic 2,3,7,8-tetrachlorodibenzo-*p*-dioxin treatment. *Cancer Res.* **60**, 5414–5419 (2000).
10. Sogawa, K. & Fujii-Kuriyama, Y. Ah receptor, a novel ligand-activated transcription factor. *J. Biochem. (Tokyo)* **122**, 1075–1079 (1997).
11. Schmidt, J. V. & Bradfield, C. A. Ah receptor signaling pathways. *Annu. Rev. Cell Dev. Biol.* **12**, 55–89 (1996).
12. Kato, S. *et al.* Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**, 1491–1494 (1995).
13. McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474 (2002).
14. Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995).
15. Freedman, L. P. Increasing the complexity of coactivation in nuclear receptor signaling. *Cell* **97**, 5–8 (1999).
16. Watanabe, M. *et al.* A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.* **20**, 1341–1352 (2001).
17. Yanagisawa, J. *et al.* Nuclear receptor function requires a TIFC-type histone acetyl transferase complex. *Mol. Cell* **9**, 553–562 (2002).
18. Kamei, Y. *et al.* A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* **85**, 403–414 (1996).
19. Tian, Y., Ke, S., Thomas, T., Meeker, R. J. & Gallo, M. A. Transcriptional suppression of estrogen receptor gene expression by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). *J. Steroid Biochem. Mol. Biol.* **67**, 17–24 (1998).
20. Tora, L. *et al.* The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* **59**, 477–487 (1989).
21. Kobayashi, Y. *et al.* p300 mediates functional synergism between AF-1 and AF-2 of estrogen receptor alpha and beta by interacting directly with the N-terminal A/B domains. *J. Biol. Chem.* **275**, 15645–15651 (2000).
22. Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes Dev.* **13**, 20–25 (1999).
23. Beischlag, T. V. *et al.* Recruitment of the NCoA/SRC-1/p160 family of transcriptional coactivators by the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator complex. *Mol. Cell. Biol.* **22**, 4319–4333 (2002).
24. Daitoku, H. Y. K., Matsuzaki, H., Hata, M. & Fukamizu, A. Regulation of PGC-1 promoter activity by protein kinase B and the forkhead transcription factor FKHR. *Diabetes* **52**, 642–649 (2003).
25. Weisz, A. & Rosales, R. Identification of an estrogen response element upstream of the human *c-fos* gene that binds the estrogen receptor and the AP-1 transcription factor. *Nucleic Acids Res.* **18**, 5097–5106 (1990).
26. Mueller, M. D. *et al.* Regulation of vascular endothelial growth factor (VEGF) gene transcription by estrogen receptors alpha and beta. *Proc. Natl Acad. Sci. USA* **97**, 10972–10977 (2000).
27. Puga, A., Nebert, D. W. & Carrier, F. Dioxin induces expression of *c-fos* and *c-jun* proto-oncogenes and a large increase in transcription factor AP-1. *DNA Cell Biol.* **11**, 269–281 (1992).
28. Mimura, J. *et al.* Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes Cells* **2**, 645–654 (1997).
29. Dupont, S. *et al.* Effect of single and compound knockouts of estrogen receptors α (ER α) and β (ER β) on mouse reproductive phenotypes. *Development* **127**, 4277–4291 (2000).
30. Kuchenhoff, A. *et al.* Arylhydrocarbon receptor expression in the human endometrium. *Fertil. Steril.* **71**, 354–360 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Korach and A. Fukamizu for helpful discussion; T. Sato, A. Murayama and Y. Kobayashi for technical assistance; Taiho Pharmaceutical Co. for ER ligands; and R. Nakamura and H. Higuchi for manuscript preparation. This work was supported in part by grants-in-aid for priority areas from the Ministry of Education, Science, Sports and Culture of Japan (to Y.F.-K. and S.K.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and request s for materials should be addressed to S.K. (uskato@mail.ecc.u-tokyo.ac.jp).

Insulin-regulated hepatic gluconeogenesis through FOXO1–PGC-1 α interaction

Pere Puigserver^{*†}, James Rhee^{*}, Jerry Donovan^{*}, Christopher J. Walkey^{*}, J. Cliff Yoon^{*}, Francesco Oriente[‡], Yukari Kitamura[‡], Jennifer Altomonte[§], Hengjiang Dong[§], Domenico Accili[‡] & Bruce M. Spiegelman^{*}

^{*} Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡] Naomi Berrie Diabetes Center and Department of Medicine, College of Physicians & Surgeons of Columbia University, New York, New York 10032, USA
[§] Institute for Human Gene Therapy and Molecular Medicine, Mount Sinai School of Medicine, New York, New York 10029, USA

Hepatic gluconeogenesis is absolutely required for survival during prolonged fasting or starvation, but is inappropriately activated in diabetes mellitus. Glucocorticoids and glucagon have strong gluconeogenic actions on the liver. In contrast, insulin suppresses hepatic gluconeogenesis^{1–3}. Two components known to have important physiological roles in this process are the forkhead transcription factor FOXO1 (also known as FKHR) and peroxisome proliferative activated receptor- γ co-activator 1 (PGC-1 α ; also known as PPARGC1), a transcriptional co-activator; whether and how these factors collaborate has not been clear. Using wild-type and mutant alleles of FOXO1, here we show that PGC-1 α binds and co-activates FOXO1 in a manner inhibited by Akt-mediated phosphorylation. Furthermore, FOXO1 function is required for the robust activation of gluconeogenic gene expression in hepatic cells and in mouse liver by PGC-1 α . Insulin suppresses gluconeogenesis stimulated by PGC-1 α but co-expression of a mutant allele of FOXO1 insensitive to insulin completely reverses this suppression in hepatocytes or transgenic mice. We conclude that FOXO1 and PGC-1 α interact in the execution of a programme of powerful, insulin-regulated gluconeogenesis.

Two transcriptional components that are targets of insulin signalling, and that can activate the process of gluconeogenesis in liver, are FOXO1 and PGC-1 α . FOXO1 has been shown to bind directly to the promoters of gluconeogenic genes and activate the process of glucose production^{4–6}. FOXO1 is directly phosphorylated by Akt, a key protein kinase downstream of the insulin receptor^{7,8}. This phosphorylation results in exclusion of FOXO1 from the nucleus. A second transcriptional component controlled by insulin and having a role in gluconeogenesis is the co-activator PGC-1 α . PGC-1 α is induced in liver on fasting, and is elevated in several models of diabetes or deficiency in insulin signalling. Notably, expression of PGC-1 α at physiological levels turns on the entire programme of gluconeogenesis^{9,10}.

[†] Present address: Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Insulin is a dominant suppressor of gluconeogenesis, and several animal models of deficiency in insulin signalling show a rise in hepatic PGC-1 α expression⁹. As insulin may control PGC-1 α expression *in vivo* either through direct action on hepatocytes or through counter-regulatory hormones such as glucagon, we investigated this question using isolated hepatocytes. As shown in Fig. 1a, b, and as expected^{9,11}, PGC-1 α messenger RNA (*Pgc1*) levels were increased by forskolin and/or dexamethasone in both immortalized and in primary mouse hepatocytes. Notably, insulin treatment does not change the level of *Pgc1* mRNA when combined with forskolin alone, or with forskolin and dexamethasone. PGC-1 α protein expression was also induced under the forskolin and dexamethasone treatment, but was unaffected by insulin (Fig. 1a). We also analysed the direct effects of insulin on *Pgc1* mRNA in mice subjected to hyperinsulinaemic, euglycaemic clamps. This procedure elevates insulin levels while maintaining steady glucose concentrations, thus preventing the rise of counter-regulatory hormones such as glucagon. As shown in Fig. 1c, elevated insulin concentrations reduced phosphoenolpyruvate carboxykinase 1 (PEPCK) mRNA (*Pck1*) but did not alter *Pgc1* mRNA levels. Together, these data indicate that insulin does not have a direct effect on PGC-1 α expression in cultured hepatocytes or liver; the effect of this hormone on PGC-1 α expression *in vivo* is probably controlled, at least in part, thorough counter-regulatory hormones, especially glucagon.

To investigate how insulin regulates the function of PGC-1 α , we used adenoviruses expressing this co-activator in Fao hepatoma cells. These cells express very little endogenous PGC-1 α and thus represent an excellent system in which to study the function of this expressed protein. As shown in Fig. 1d, and as expected⁹, expression of PGC-1 α strongly induced the gluconeogenic genes *Pck1* and glucose-6-phosphatase (*G6pc*). The addition of insulin at 10 nM greatly reduced the amount of these mRNAs driven by PGC-1 α . However, mitochondrial genes that are targets of PGC-1 α , such as cytochrome *c* and the β -subunit of ATP synthase, were not regulated by the insulin treatment. These results demonstrate that insulin acts directly on hepatocytes to suppress the regulatory function of PGC-1 α in the expression of a certain subset of genes, especially those encoding the gluconeogenic enzymes.

As the cellular distribution of FOXO1 is regulated by Akt in response to insulin^{8,12}, this factor is considered to be a potential key mediator of the insulin repression of the gluconeogenic genes. To investigate whether PGC-1 α requires FOXO1 to induce gluconeogenic genes, we used adenoviral expression of FOXO1(1–256), a truncated allele of FOXO1 that lacks the transactivation domain and functions as a dominant-negative suppressor of wild-type FOXO1 (ref. 13). Whereas PGC-1 α robustly activated expression of *G6pc* and *Pck1* (Fig. 2a), FOXO1(1–256) had no ability to activate *G6pc*, and had only a relatively weak effect on *Pck1*. However, co-expression of FOXO1(1–256) markedly diminished the ability of

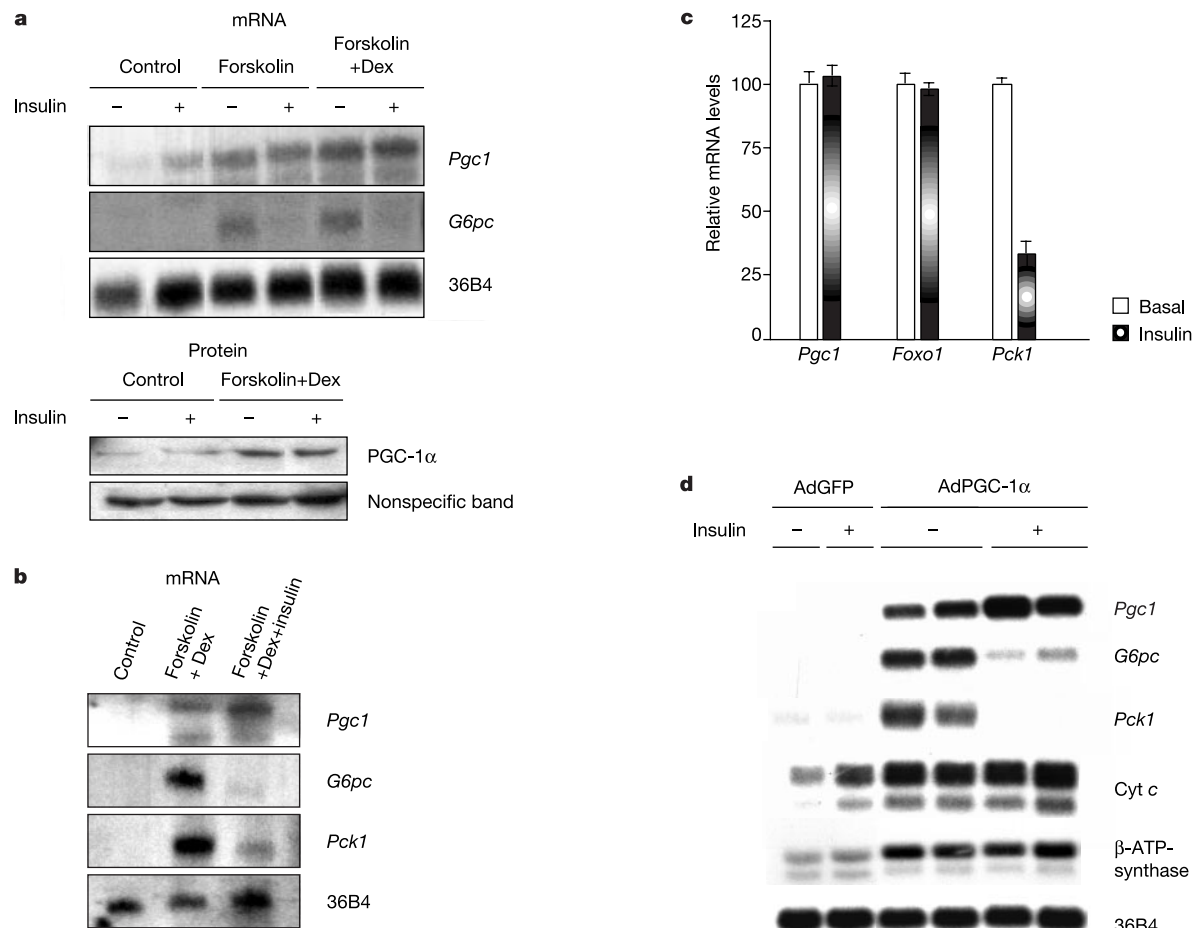


Figure 1 Insulin regulates PGC-1 α action but not PGC-1 α gene expression. **a**, PGC-1 α mRNA and protein amounts are not regulated by insulin in mouse hepatocytes. Immortalized hepatocytes were differentiated as described in Methods. **b**, *Pgc1* mRNA amounts are not changed by insulin in primary hepatocytes. **c**, *Pgc1* mRNA amounts are not changed by insulin in mice. Mice were subjected to glucose clamps as described in

the Methods. Three mice for each group were used. **d**, Insulin markedly decreases PGC-1 α induction of gluconeogenic genes. Fao rat hepatocytes were placed in serum-free medium and infected with adenovirus vectors encoding either GFP or PGC-1 α . Dex, dexamethasone.

PGC-1 α to activate these gluconeogenic genes. Notably, FOXO1 (1–256) had no effect on the ability of PGC-1 α to increase expression of mRNAs for the mitochondrial proteins cytochrome *c* or β -ATP synthase. These data indicate that PGC-1 α requires intact FOXO1 signalling to activate the gluconeogenic genes, but not for certain other functions. These data also suggest that PGC-1 α and FOXO1 functionally interact.

A functional FOXO1–PGC-1 α interaction was studied more directly using immortalized hepatocytes transfected with a consensus response sequence for FOXO1 together with a PGC-1 α expression plasmid or a control vector. As shown in Fig. 2b, co-expression of PGC-1 α increased the transcriptional activity of wild-type FOXO1 fourfold. Similar results were obtained when PGC-1 α was co-expressed with a constitutively active mutant of FOXO1 (FOXO1(3A)) in which the three sites for phosphorylation by Akt have been substituted to alanine⁷. The addition of insulin repressed the PGC-1 α activation of wild-type FOXO1 but did not suppress the effect of PGC-1 α on FOXO1(3A). Similar results were obtained using the promoter for glucose-6-phosphatase (Supplementary Fig. S1). Taken together these results show that PGC-1 α augments the tran-

scription of FOXO1 in an insulin-dependent way, and that a constitutively active form of FOXO1 completely ablates the repressive effect that insulin has over the action of PGC-1 α on a gluconeogenic gene promoter. We next analysed whether the PGC-1 α -FOXO1 interaction involved direct physical binding. As shown in Fig. 2c, PGC-1 α and FOXO1 formed a specific precipitable complex. To investigate the protein domains involved in this interaction, recombinant forms of these two proteins were constructed and tested for *in vitro* interactions. FOXO1 interacted with the carboxy-terminal part of PGC-1 α . PGC-1 α binds to the amino-terminal part of FOXO1 (Fig. 2d; see also Supplementary Fig. S2). These data illustrate that PGC-1 α and FOXO1 directly interact, indicating that PGC-1 α acts as a direct transcriptional co-activator of FOXO1. To investigate whether insulin also disrupts the binding of PGC-1 α to the endogenous PEPCK and glucose-6-phosphatase promoters, we performed chromatin immunoprecipitation. As shown in Fig. 2e, endogenous PGC-1 α protein was recruited to FOXO1 binding sites of the PEPCK promoter (AF2) and the glucose-6-phosphatase promoter (IRU)^{4,14}. In both cases, this binding was completely abolished by insulin treatment. These data indicate that PGC-1 α -mediated expression of

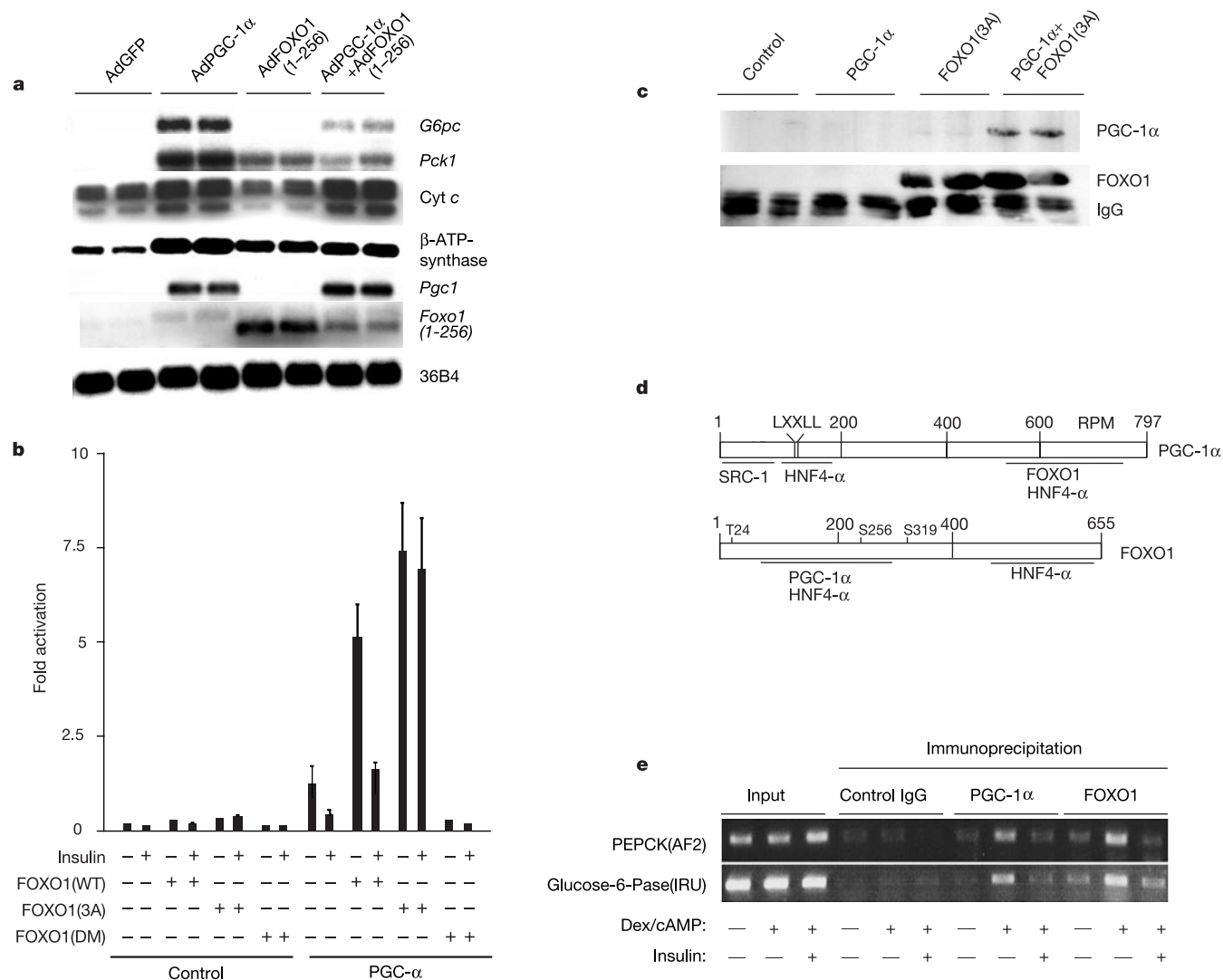


Figure 2 Co-activation of FOXO1 by PGC-1 α is required for gluconeogenic gene expression. **a**, PGC-1 α induction of gluconeogenic genes is blocked by a dominant-negative effect of FOXO1. Fao hepatocytes were infected as described in Methods. **b**, PGC-1 α co-activates FOXO1. Immortalized hepatocytes were transfected with an insulin response element/luciferase reporter gene and the plasmids indicated. DM, DNA-binding mutant. **c**, PGC-1 α interacts with FOXO1. BOSC cells were transfected with vector,

PGC-1 α and Flag-tagged FOXO1(3A) plasmids. Immunoprecipitation was performed as in Methods. **d**, Interaction domains in PGC-1 α and FOXO1. RPM, RNA-processing motifs. **e**, Insulin decreases endogenous PGC-1 α binding to the PEPCk and glucose-6-phosphatase promoters containing FOXO1 response elements. ChIP assays were performed as described in Methods.

gluconeogenic genes involves the binding of endogenous PGC-1 α to FOXO1 response elements in chromatin.

Whether the functional interaction between PGC-1 α and FOXO1 might be responsible for the ability of insulin to suppress the PGC-1 α -mediated activation of endogenous gluconeogenic genes was investigated using the constitutively active allele of FOXO1. As shown in Fig. 3a, insulin can suppress PGC-1 α -mediated induction of the gluconeogenic genes when this co-activator is expressed alone, or when it is co-expressed with wild-type FOXO1. However, when PGC-1 α is co-expressed with the constitutively active FOXO1 allele, the expression of glucose-6-phosphatase and PEPCK is no longer sensitive to insulin. Adenoviral-mediated expression of FOXO1 proteins without PGC-1 α has little effect on the gluconeogenic genes (data not shown). We next investigated whether these changes in transcriptional activity and induction of gluconeogenic genes were reflected in glucose production *per se*. As shown in Supplementary Fig. S3, adenoviral-mediated PGC-1 α expression increased glucose secretion threefold compared with cells infected with control virus. Consistent with the induction of the gluconeogenic genes shown above, insulin repressed the glucose production by PGC-1 α , but this repression was overcome by co-expression of constitutively active FOXO1.

To investigate whether insulin prevents the physical interaction between PGC-1 α and FOXO1, we analysed this complex in cells by co-immunoprecipitation after transfection of cells. As shown in Fig. 3b, a constitutively active form of Akt, which acts downstream of insulin and phosphorylates FOXO1, clearly interferes with the interaction between PGC-1 α and FOXO1. FOXO1(3A) interacted with PGC-1 α independently of Akt expression. To determine whether phosphorylation of FOXO1 by Akt also disrupts the PGC-1 α -FOXO1 interaction *in vitro*, we performed binding with unmodified and phosphorylated FOXO1 and the glutathione

S-transferase (GST)-PGC-1 α fusion protein. This experiment was done in two ways: using Akt-mediated phosphorylation of FOXO1 before binding to PGC-1 α (Fig. 3c) or after binding to PGC-1 α (Fig. 3d). As shown in Fig. 3c, phosphorylation of FOXO1 by Akt strongly reduces its binding to PGC-1 α ($62 \pm 5\%$); as a control, the binding of FOXO1(3A) was not affected in the presence of Akt and ATP. Markedly similar results were observed when the PGC-1 α -FOXO1 complex was preformed. The PGC-1 α -FOXO1 complex was largely disrupted by Akt ($60 \pm 7\%$), and again the mutant FOXO1(3A) binding was not affected (Fig. 3d). These results strongly indicate that, in addition to causing exclusion from the nucleus, phosphorylation of FOXO1 by Akt specifically disrupts the interaction with PGC-1 α .

We next investigated whether a functional interaction between FOXO1 and PGC-1 α occurs in live animals, by infusing recombinant adenoviruses expressing PGC-1 α or the dominant-negative version of FOXO1 into mice. As expected, PGC-1 α increased the levels of gluconeogenic genes *Pck1* and *G6pc* in infected mice. Notably, co-infection with the dominant-negative allele of FOXO1 reduced the levels of these PGC-1 α -induced gluconeogenic genes, indicating that activation of gluconeogenic genes by PGC-1 α requires FOXO1 function (Fig. 4a). To evaluate whether insulin regulation of PGC-1 α function requires the FOXO1-PGC-1 α interaction *in vivo*, adenoviruses expressing PGC-1 α or control green fluorescent protein (GFP) were infused into wild-type and transgenic mice expressing the constitutively active form of FOXO1 in liver. These transgenic mice are slightly hyperglycaemic and demonstrate a moderate induction of gluconeogenic genes⁶. As expected, PGC-1 α induced an increase of *G6pc* and *Pck1* mRNAs in the liver of wild-type animals⁹, and caused a similar response in transgenic mice (Fig. 4b). However, whereas injection of insulin decreased the PGC-1 α -mediated expression of *Pck1* and *G6pc*

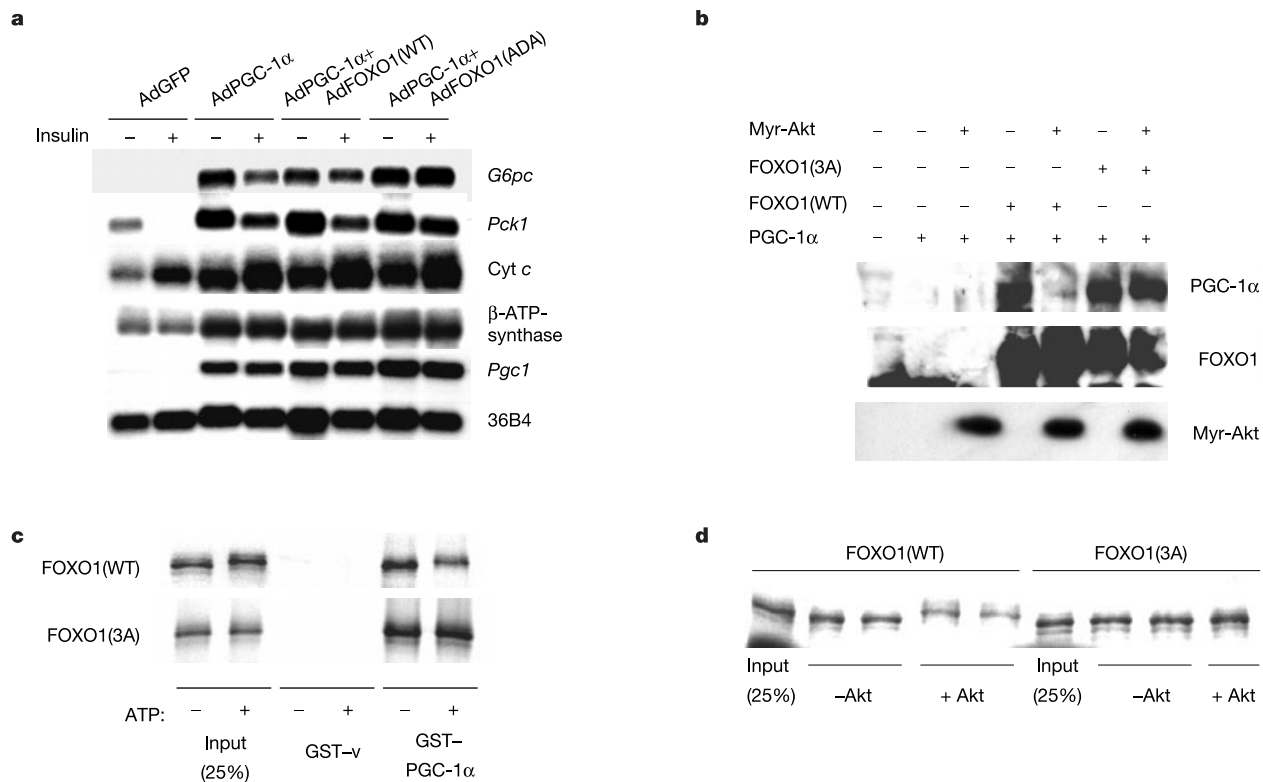


Figure 3 Insulin regulation of PGC-1 α action on gluconeogenesis depends on FOXO1 function in hepatocyte cells. **a**, Insulin regulation of PGC-1 α -mediated gluconeogenic gene expression is abolished by a constitutively active form of FOXO1. Fao rat hepatocytes were treated as described in Fig. 2. **b**, Constitutively activated Akt disrupts PGC-1 α -FOXO1 interaction. BOSC cells were transfected with the indicated plasmids and co-

immunoprecipitation was performed as described in the Methods. **c**, Phosphorylation of FOXO1 by Akt decreases binding to PGC-1 α . Binding assays were performed as described in Supplementary Fig. S2. **d**, Pre-bound PGC-1 α -FOXO1 complex is largely disrupted by Akt-mediated phosphorylation of FOXO1. Similar results were obtained in three different experiments.

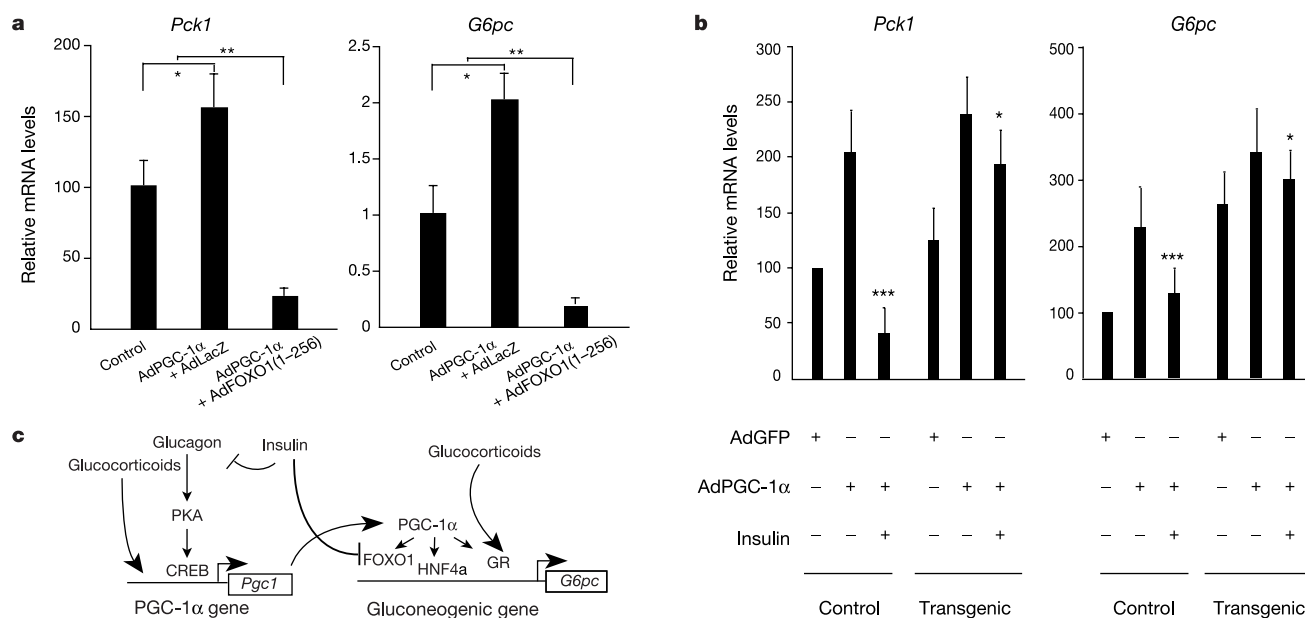


Figure 4 Insulin regulation of PGC-1 α action on gluconeogenic genes depends on FOXO1 function in live mice. **a**, Dominant-negative FOXO1 (FOXO1(1–256)) suppresses PGC-1 α induction of gluconeogenic genes. Mice were injected with adenoviruses expressing control LacZ, PGC-1 α and FOXO1(1–256) as described in the Methods. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$. **b**, Insulin does not suppress PGC-1 α -induced

gluconeogenic genes in FOXO1(S253A) transgenic mice. Mice were injected with adenoviruses expressing GFP or PGC-1 α as described in the Methods. Asterisk, $P < 0.01$ compared with wild-type mice; triple asterisk, $P < 0.05$ compared with fasting values in PGC-1 α -expressing wild-type mice. **c**, Model illustrating the influence of the major hormones on the gluconeogenic genes. GR, glucocorticoid receptor.

mRNAs in control mice by 80% and 44%, respectively, a much smaller decrease in these mRNAs in response to insulin was observed in the constitutive FOXO1 transgenic mice (12% and 18%, respectively). Amounts of adenoviral-mediated PGC-1 α and FOXO proteins in mice were equivalent (data not shown). Taken together, these data demonstrate that the ability of PGC-1 α to induce gluconeogenesis requires its interaction with FOXO1, and that the ability of insulin to suppress PGC-1 α -mediated gluconeogenesis, both in cells and in live animals, is dependent on the ability of the insulin pathway to modify FOXO1 activity.

These data, combined with previous studies, allows a model that can readily incorporate all the classical hormonal influences on the process of gluconeogenesis (Fig. 4c). Glucagon and glucocorticoids are elevated in fasting and activate expression of *Pgc1*. In addition, PGC-1 α co-activates the liganded glucocorticoid receptor^{9,15}. Insulin has two distinct modes of action here: the lack of any suppression of PGC-1 α expression by insulin in isolated hepatocytes and in mice under hyperinsulinaemic, euglycaemic clamp conditions suggests that the elevation of PGC-1 α observed in various states of insulin deficiency may occur, at least in part, through its well-known suppression of glucagon release. However, as it is impossible to mimic the exact physiological environment in cell culture, a direct effect of insulin on hepatic PGC-1 α expression cannot be ruled out. Another study has shown a transient effect of insulin on PGC-1 α expression¹⁰, although we have not been able to observe this. What can be readily observed here, however, is a direct suppressive effect of insulin downstream of PGC-1 α . This effect seems to be through modification of FOXO1, in that expression of an allele of FOXO1 that is not modified by insulin/Akt signalling renders the gluconeogenic function of PGC-1 α completely insensitive to insulin. As demonstrated previously⁷, phosphorylation of FOXO1 by Akt causes its sequestration in the cytoplasm. As we show here, this modification of FOXO1 also has another function; that is, it specifically disrupts the FOXO1–PGC-1 α interaction.

Although physical interaction between PGC-1 α and other transcription factors such as the nuclear receptors HNF4- α and glucocorticoid receptor are probably critically important for this

response, FOXO1 is the first transcription factor shown to be required for the gluconeogenic action of PGC-1 α . Hence the PGC-1 α –FOXO1 complex must be considered a potential target for anti-gluconeogenic therapies for diabetes mellitus. The fact that this complex can be disrupted by a small number of phosphorylations raises hope that a small molecule can be developed that also inhibits this interaction. Although the model shown in Fig. 4c can integrate several of the known gluconeogenic hormones and transcription factors as regulators of PGC-1 α , it is worth noting that certain others do not yet fall easily into this scheme. The transcription factors C/EBP- α and - β as well as ADD1/SREBP1 are highly enriched in liver and have been shown to regulate hepatic gluconeogenesis^{16–18}. Whether PGC-1 α also co-activates these transcription factors or is a target of these factors is under investigation. □

Methods

Cell culture and treatments

Primary mouse hepatocytes were isolated and cultured in DMEM with 10% fetal bovine serum (FBS) as described¹⁹. Treatment with insulin (10 nM) and forskolin (1 μ M) was performed in 0.5% bovine serum albumin (BSA) for 14 h. Murine hepatocytes immortalized by SV40 large T antigen¹³ were cultured in α -MEM with 100 nM dexamethasone and 4% fetal calf serum. Fao rat hepatoma cells were cultured in RPMI medium with 10% fetal calf serum.

Transcriptional activation assays

Immortalized mouse hepatocytes were transiently transfected using FuGENE (Roche) or Superfect (Qiagen). After overnight incubation, medium was changed to 0.5% BSA in DMEM medium for 24 h. Insulin was added 2 h before the addition of dexamethasone and forskolin, which were present for the last 6 h. Cells were lysed and aliquots were used to measure β -galactosidase and luciferase activities. pcDNA3-Flag-FOXO1 plasmids were a gift from W. Sellers. PGC-1 α plasmids have been described previously²⁰.

Protein interaction analysis

pGEX2-PGC-1 α and pGEX2-FOXO1 plasmids were generated by cloning the corresponding polymerase chain reaction (PCR) fragments into the *Bam*H1 and *Xho*I cloning sites of these vectors. GST–PGC-1 α and GST–FOXO1 fragments were expressed in bacteria (BL21) by isopropyl- β -D-thiogalactoside induction for 3 h at room temperature. Fusion proteins were purified on Sepharose beads containing glutathione. [³⁵S]-labelled proteins were made with a TNT reticulocyte lysate kit (Promega). Equal amounts of GST fusion proteins (1 μ g) were mixed with 5 μ l of the *in vitro* translated proteins in a binding buffer containing 20 mM HEPES buffer (pH 7.7), 75 mM KCl, 0.1 mM EDTA, 2.5 mM

MgCl₂, 0.05% NP40, 2 mM dithiothreitol and 10% glycerol. Binding reactions were performed as in ref. 20.

For protein interaction experiments involving Akt-mediated phosphorylation, binding was performed as described above, and *in vitro* translated FOXO1 was phosphorylated with activated Akt (Upstate) following the manufacturer's instructions.

Co-immunoprecipitation experiments

Flag-tagged FOXO1 and PGC-1 α proteins were expressed in BOSC23 cells using FuGENE (Roche). Forty-eight hours after transfection, whole-cell extracts were prepared and subjected to an overnight incubation with a monoclonal antibody to Flag (Sigma) linked to agarose beads. The immunoprecipitates were washed four times with lysis buffer, separated by SDS-polyacrylamide gel electrophoresis, and immunoblotted using antibodies directed against the N terminus of PGC-1 α ²⁰.

ChIP assays

Immortalized hepatocytes were fixed with formaldehyde, lysed and then sonicated. Soluble chromatin was co-immunoprecipitated with anti-FOXO1 antiserum, anti-PGC-1 α antiserum or an equal amount of immunoglobulin- γ (IgG). After de-crosslinking of the DNA, samples were subjected to PCR using the following primers: PEPCK (–438 to –326), 5'-GTGGGAGTGACACCTCACAGC-3' and 5'-AGGGCAGGCCTAGCCGAGACG-3'; glucose-6-phosphatase (–251 to –31), 5'-GCCTCTAGCACTCAAGCAGTG-3' and 5'-TGTGCCTTGCCCTGTTTATATG-3'. These regions of amplification contain the FOXO1 binding sites of both the PEPCK promoter (accessory factor 2 site; AF2) and the glucose-6-phosphatase promoter (insulin response unit; IRU). We used standard reaction conditions and 25 cycles of amplification²¹.

Adenoviral infections

Fao hepatocytes and mouse primary hepatocytes were infected with adenoviruses expressing GFP or PGC-1 α at a multiplicity of infection of approximately 50 (ref. 9), and/or FOXO1 wild type, FOXO1(ADA) (where ADA is T24A/S253D/S316A) and FOXO1(1–258) at a multiplicity of infection of approximately 25 (ref. 13). Adenoviral infection was performed in RPMI 1640 medium and 0.5% BSA. After two days cells were treated with 10 nM insulin for 12 h. Cells were collected for RNA isolation using the Trizol reagent (Invitrogen).

Animal experiments

Recombinant adenovirus encoding GFP, PGC-1 α , LacZ and dominant-negative FOXO1 were purified by CsCl gradient centrifugation and concentrated to 1.3×10^{10} plaque-forming units per ml, corresponding to 2.5×10^{12} viral particles (vp) per ml. For the PGC-1 α and dominant-negative FOXO1 experiment (Fig. 4a), ten-week-old CD-1 mice were injected with 0.75×10^{11} vp per mouse. Each animal received the same total of viral particles per body weight and LacZ virus was used as a control. At day 2, blood glucose and insulin was measured and the mice were killed, the livers were removed and analysed for mRNA isolation, and real-time PCR with reverse transcription (RT) was performed. For the PGC-1 α adenoviral infection in FOXO1 transgenic mice (Fig. 4b), six-month-old Ttr/FOXO1(S253A) male transgenic mice and wild-type littermate controls⁶ were injected with purified virus through the tail vein at a dose of 4×10^9 vp g^{–1} in a total volume of 0.1 ml. The total viral load was approximately 1.1×10^{11} per mouse. Blood was taken immediately before and five days after injection to measure plasma glucose and hepatic enzymes ALS (acid-labile subunit) and ALT (alanine-aminotransferase). At the end of the fifth day, mice fed *ad libitum* were injected with insulin (0.75 U per kg body weight) intraperitoneally. After 1 h, plasma glucose levels were measured, animals were killed, and the livers removed and frozen in liquid nitrogen. Total mRNA isolation and real-time RT-PCR was performed⁶. For the glucose clamp experiments, 6–8-week-old mice were subjected to hyperinsulinaemic, euglycaemic clamps. After a 60-min basal period, mice underwent a 90 min euglycaemic clamp period, during which insulin was infused at 18 mU kg^{–1} min^{–1} (ref. 22). At the end of the insulin infusion period, livers were removed and snap-frozen. Liver mRNA was isolated and used for real-time RT-PCR.

Received 8 January; accepted 24 April 2003; doi:10.1038/nature01667.

Published online 18 May 2003.

1. Unger, R. H. Glucagon physiology and pathophysiology in the light of new advances. *Diabetologia* **28**, 574–578 (1985).
2. Hanson, R. W. & Reshef, L. Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. *Annu. Rev. Biochem.* **66**, 581–611 (1997).
3. Hall, R. K. & Granner, D. K. Insulin regulates expression of metabolic genes through divergent signaling pathways. *J. Basic Clin. Physiol. Pharmacol.* **10**, 119–133 (1999).
4. Hall, R. K. *et al.* Regulation of phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein-1 gene expression by insulin. The role of winged helix/forkhead proteins. *J. Biol. Chem.* **275**, 30169–30175 (2000).
5. Schmolli, D. *et al.* Regulation of glucose-6-phosphatase gene expression by protein kinase B α and the forkhead transcription factor FKHR. Evidence for insulin response unit-dependent and -independent effects of insulin on promoter activity. *J. Biol. Chem.* **275**, 36324–36333 (2000).
6. Nakae, J. *et al.* Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. *Nature Genet.* **32**, 245–253 (2002).
7. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
8. Nakae, J., Park, B. C. & Accili, D. Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a Wortmannin-sensitive pathway. *J. Biol. Chem.* **274**, 15982–15985 (1999).
9. Yoon, J. C. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).

10. Herzig, S. *et al.* CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature* **413**, 179–183 (2001).
11. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
12. Rena, G., Guo, S., Cichy, S. C., Unterman, T. G. & Cohen, P. Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. *J. Biol. Chem.* **274**, 17179–17183 (1999).
13. Nakae, J., Kitamura, T., Silver, D. L. & Accili, D. The forkhead transcription factor Foxo1 (Fkhr) confers insulin sensitivity onto glucose-6-phosphatase expression. *J. Clin. Invest.* **108**, 1359–1367 (2001).
14. Durham, S. K. *et al.* FKHR binds the insulin response element in the insulin-like growth factor binding protein-1 promoter. *Endocrinology* **140**, 3140–3146 (1999).
15. Knutti, D., Kaul, A. & Kralli, A. A tissue-specific coactivator of steroid receptors, identified in a functional genetic screen. *Mol. Cell. Biol.* **20**, 2411–2422 (2000).
16. Darlington, G. J., Wang, N. & Hanson, R. W. C/EBP α : a critical regulator of genes governing integrative metabolic processes. *Curr. Opin. Genet. Dev.* **5**, 565–570 (1995).
17. Arizmendi, C., Liu, S., Croniger, C., Poli, V. & Friedman, J. E. The transcription factor CCAAT/enhancer-binding protein β regulates gluconeogenesis and phosphoenolpyruvate carboxykinase (GTP) gene transcription during diabetes. *J. Biol. Chem.* **274**, 13033–13040 (1999).
18. Becard, D. *et al.* Adenovirus-mediated overexpression of sterol regulatory element binding protein-1c mimics insulin effects on hepatic gene expression and glucose homeostasis in diabetic mice. *Diabetes* **50**, 2425–2430 (2001).
19. Zaher, H. *et al.* The involvement of aryl hydrocarbon receptor in the activation of transforming growth factor- β and apoptosis. *Mol. Pharmacol.* **54**, 313–321 (1998).
20. Puigserver, P. *et al.* Activation of PPAR γ coactivator-1 through transcription factor docking. *Science* **286**, 1368–1371 (1999).
21. Nakae, J. *et al.* The forkhead transcription factor Foxo1 regulates adipocyte differentiation. *Dev. Cell* **4**, 119–129 (2003).
22. Rossetti, L. *et al.* Hepatic overexpression of insulin-like growth factor-II in adulthood increases basal and insulin-stimulated glucose disposal in conscious mice. *J. Biol. Chem.* **271**, 203–208 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge members of the Spiegelman laboratory for helpful discussions on the project. We also thank P. Vazquez for discussions on the project. Some constructs and reagents used in this work were obtained from W. Sellers, D. Schmolli, Y. Inoue and F. Gonzalez. We also thank the Vector Core of the Institute for Human Gene Therapy at Mount Sinai School of Medicine for virus preparation and help with injections. P.P. was supported by a Lee Career Award. This work was supported by grants to B.M.S. from the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.M.S. (bruce_spiegelman@dfci.harvard.edu).

Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors

Zhulun Wang^{*†‡}, Gérard Benoit^{†‡}, Jinsong Liu^{*}, Srividya Prasad^{*}, Piia Aarnisalo[‡], Xiaohong Liu^{*}, Haoda Xu^{*}, Nigel P. C. Walker^{*} & Thomas Perlmann^{‡§}

^{*} Department of Structural Biology, Tularik Inc., 1120 Veterans Blvd., South San Francisco, California 94080, USA

[‡] Ludwig Institute for Cancer Research, Box 240, § Department of Cell and Molecular Biology, Karolinska Institutet, S-171 77 Stockholm, Sweden

[†] These authors contributed equally to this work

Members of the nuclear receptor (NR) superfamily of transcription factors modulate gene transcription in response to small lipophilic molecules¹. Transcriptional activity is regulated by ligands binding to the carboxy-terminal ligand-binding domains (LBDs) of cognate NRs. A subgroup of NRs referred to as 'orphan receptors' lack identified ligands, however, raising issues about the function of their LBDs². Here we report the crystal structure of the LBD of the orphan receptor Nurr1 at 2.2 Å resolution. The Nurr1 LBD adopts a canonical protein fold resembling that of

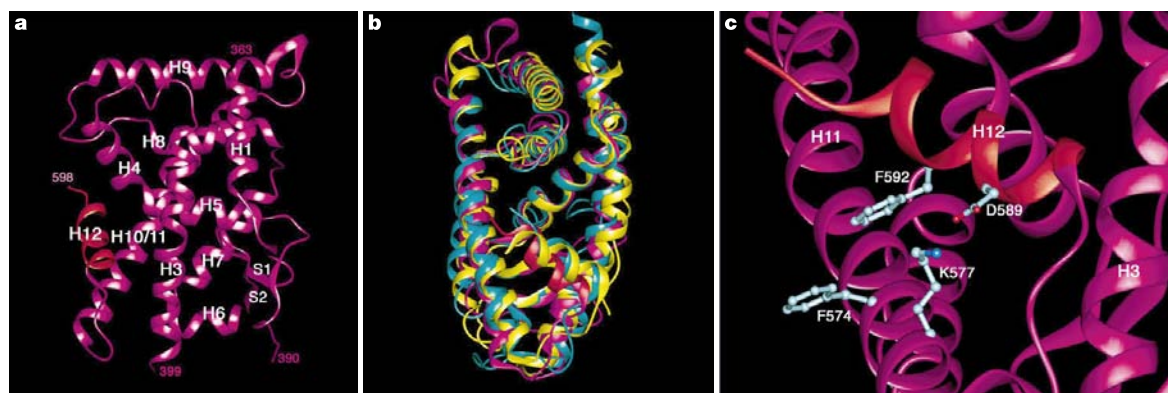


Figure 1 Ribbon representations of NR LBD structures. **a**, Nurr1 LBD contains a three-layered antiparallel α -helical sandwich formed by 12 α -helices (H1–H12) and one β -sheet of two strands (S1, S2). The AF2 helix (H12) is shown in red. **b**, Superposition of the Nurr1 LBD in magenta, with holo RAR γ (PDB code: 2LBD) in cyan and holo ER α (PDB

code: 1ERE) in yellow. The view is rotated by 90° relative to **a**. **c**, Interactions between H11 and H12 that position and stabilize H12 in the transcriptionally active conformation in Nurr1 LBD. F592 from H12 forms an aromatic stacking interaction with F574 from H11, whereas K577 from H11 forms a salt-bridge with D589 from H12.

agonist-bound, transcriptionally active LBDs in NRs³, but the structure has two distinctive features. First, the Nurr1 LBD contains no cavity as a result of the tight packing of side chains from several bulky hydrophobic residues in the region normally occupied by ligands. Second, Nurr1 lacks a ‘classical’ binding site for coactivators. Despite these differences, the Nurr1 LBD can be regulated in mammalian cells. Notably, transcriptional activity is correlated with the Nurr1 LBD adopting a more stable conformation. Our findings highlight a unique structural class of NRs and define a model for ligand-independent NR function.

Roughly half of the mammalian NRs and most invertebrate NRs are currently classified as orphans because their *bona fide* endogenous ligands have not been identified. The search for ligands for orphan NRs has become the subject of intense investigation and debate. The successful identification of fatty acids, oxysterols and bile acids as naturally occurring agonists of the peroxisome proliferator-activated receptors (PPARs), the liver X receptors and the farnesoid X receptor, respectively, supports the notion that all NRs are regulated by ligands^{4,5}. However, efforts to identify ligands for many other NRs have met with less success and structural studies of orphan receptors have suggested that some NRs such as hepatocyte nuclear factor 4 α (HNF4 α) and HNF4 γ , although capable of binding ligands, may not be ligand-regulated^{6,7}. But definitive evidence of a NR that is not ligand-regulated is still lacking.

Nurr1 (also known as NR4A2), an orphan NR lacking identified ligands, belongs to the nerve growth factor-induced clone B (NGFI-B) subfamily of orphan receptors that can interact with DNA as monomers or homodimers. Nurr1 and NGFI-B can also interact with DNA as heterodimers with retinoid X receptors (RXRs). Nurr1 is essential for the development of dopamine neurons in mice^{8–10} and is linked to cases of familial Parkinson’s disease in humans¹¹. Although no ligand for Nurr1 has been identified, its LBD shows constitutive and cell-type-specific activity that is dependent on an unusual sequence in the activation function-2 (AF2) region¹². To facilitate identification of a putative Nurr1 ligand and to characterize further the mechanisms of Nurr1 transcriptional activation, we have determined the structure of the Nurr1 LBD by X-ray crystallography.

The overall structure of the Nurr1 LBD (Fig. 1a) adopts the characteristic canonical fold of NR LBDs^{3,13}. Globally, Nurr1 LBD is most similar to the LBD of the holo retinoic acid receptor (RAR γ)¹⁴ but superimposes relatively well with other ligand-bound LBDs (Fig. 1b), such as the 17 β -oestradiol (E₂)-bound oestrogen receptor α (ER α)¹⁵. In this conformation, the AF2 helix folds back towards the body of the LBD and packs against helices H3, H4 and H10, with its hydrophobic residues protruding into the

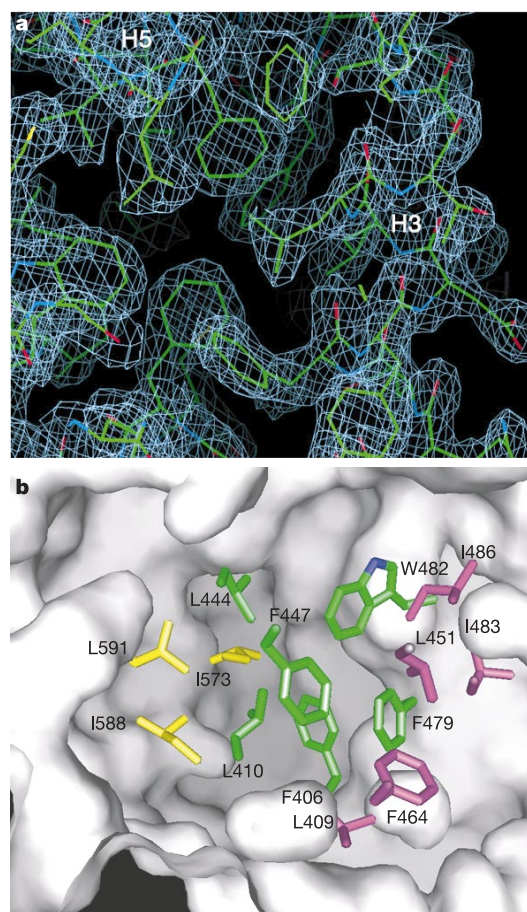


Figure 2 No ligand-binding cavity in Nurr1. **a**, The σ_A -weighted $2F_o - F_c$ electron density map (contoured at 1.5 σ) of the Nurr1 LBP behind α -helix H3. **b**, The six residues that completely block the ligand passage behind H3 (F406 and L410 from H3, L444 and F447 from H5, F479 from loop 6–7 and W482 from H7) are shown in green. The pocket is further sealed by residues L409 from H3, L451 from H5, I483 and I486 from H7 and F464 from the β -turn on one side the structure, and residues I573 from H11, and I588 and L591 from H12 on the other side, shown in pink and yellow, respectively. The internal surface of the protein, after removing these side chains, is shown in grey.

core of the LBD. Evidently, a salt bridge and several hydrophobic interactions between helices H12 and H11 are crucial for stabilizing the AF2 helix in a conformation that would otherwise be achieved only through ligand interactions in holo NRs (Fig. 1c). These structural observations explain two previously reported mutations¹², namely, that both D589A and F592A mutants abolish AF2-dependent transactivation by monomeric Nurr1 in human embryonic kidney 293 cells. Thus, the tertiary structure of the Nurr1 LBD maintains an architecture that is very similar to that of ligand-bound NR LBDs; however, the AF2 helix conformation observed in Nurr1 is stabilized by intramolecular interactions in the absence of any ligand.

The NR ligand-binding pocket (LBP) is located in the lower part of the LBD. Ligand recognition and the specificity of ligand binding are largely determined by the size and shape of this pocket, which is altered by the side chains that constitute its lining. The LBDs of NRs reported so far have a pocket of varying size, with PPAR γ having the largest cavity¹⁶ and oestrogen-receptor-related (ERR3) having the smallest¹⁷. Unexpectedly, the Nurr1 LBD structure shows that it contains no ligand-binding cavity (Fig. 2a). Instead, several tightly packed, bulky hydrophobic residues occupy the space that is available for ligand binding in other NRs. The six key hydrophobic residues that contribute to filling this space completely block the ligand passage behind helix H3 (Fig. 2b).

These findings provide evidence of an NR LBD lacking a ligand-binding cavity and thus show the existence of a ligand-independent orphan NR. The NGFI-B subfamily also includes the mammalian receptors NGFI-B (NR4A1) and Nor1 (NR4A3) and the *Drosophila* homologue DHR38 (ref. 18). The important hydrophobic residues (Fig. 2b) that fill the space occupied by ligands in other NRs are all

conserved in this subfamily. Thus, the absence of an LBP is predicted for all members of the NGFI-B subfamily, suggesting that they are all ligand-independent transcription factors.

Although the position of H12 in this ligand-free structure indicates that the protein is in a transcriptionally active conformation, a cognate surface formed by H12 and parts of H3 and H4 for cooperative interactions with coactivators is distinctly different from the surface found in other NRs. In the active conformation, a 'charge clamp' is normally formed between a glutamic acid residue from H12 and a lysine residue from H3 for the purpose of interacting with the amino and carboxy termini, respectively, of a helix in a coactivator polypeptide containing an LXXLL motif^{16,17,19}. As predicted by sequence alignment, the Nurr1 LBD shows a reversed charge clamp formed by K590 from H12 and E422 from H3 (Fig. 3). K590 is unlikely to be able to perform such a role, however, because it is engaged in a salt bridge with E440 from H4.

More strikingly, the coactivator hydrophobic binding cleft seen in other NRs is completely transformed into a charged surface. The most substantial disruption to the coactivator-binding surface results from the side-chain packing of R418, which folds directly into the shallow groove and makes van der Waals contacts with F439 from H4 at the floor of the groove (Fig. 3b, d). Thus, the coactivator interaction surface observed in other NRs is significantly different in Nurr1. These findings indicate that Nurr1 is devoid of the capability to recruit and to interact with a coactivator or corepressor through a 'classical' coactivator-binding site. Indeed, the Nurr1 LBD does not interact with several tested NR coactivators including SRC-1, ACTR, PGC-1 and CBP (as assessed by two-hybrid interaction assays and reporter gene analyses; unpublished data), and coregulatory proteins that interact with the Nurr1 LBD remain to be identified.

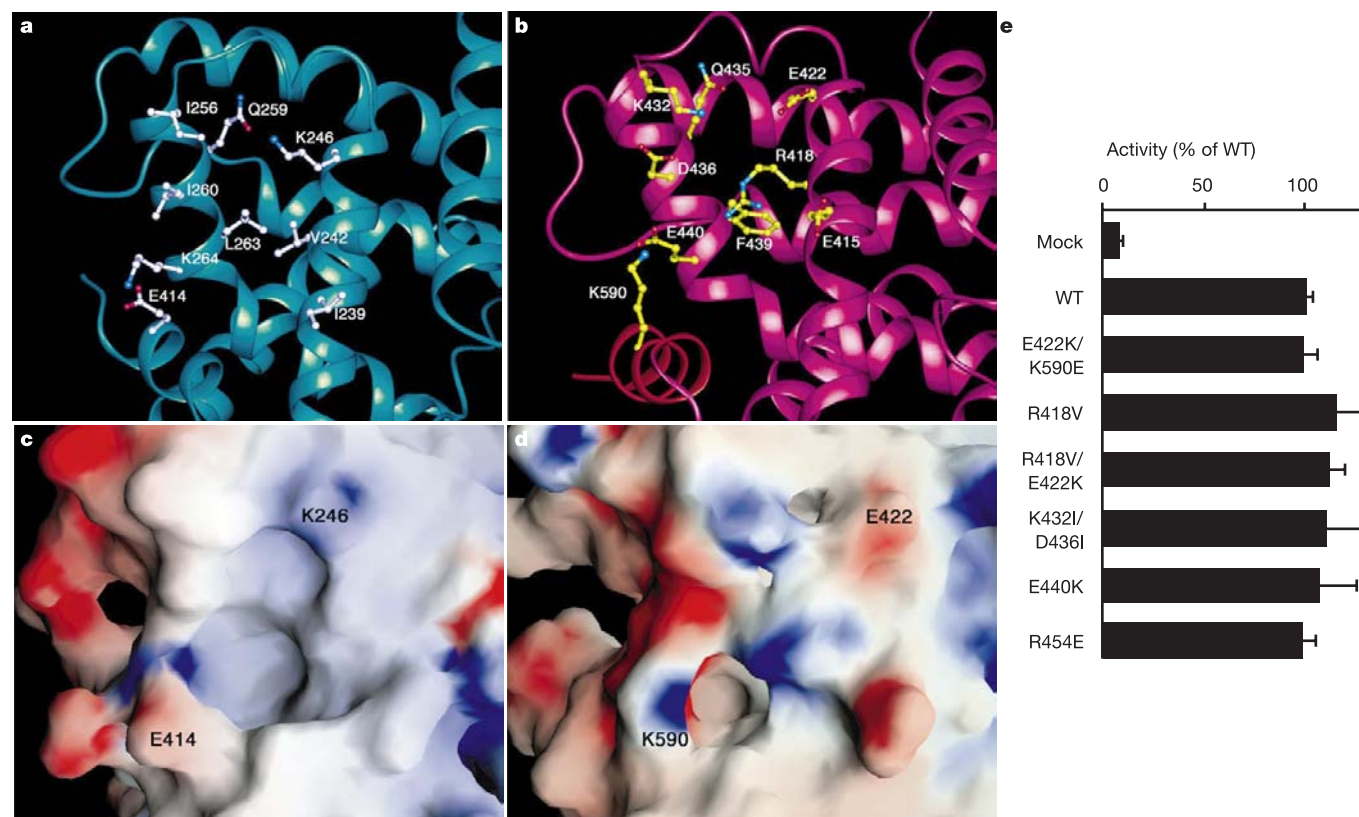


Figure 3 The region of the coactivator-binding site. **a–d**, The residues forming the coactivator interaction surface in RAR γ (**a**) are coloured white and the equivalent residues in Nurr1 (**b**) are coloured yellow. The corresponding electrostatic potential surfaces are shown in **c** and **d**, respectively. Regions of positive potential are blue, negative potential is red and neutral hydrophobic areas are white. K246 and E414 form

the 'charge clamp' residues in RAR γ . **e**, Mutations introduced to make Nurr1 resemble more the coactivator- or corepressor-interacting surface in other NRs do not change the transcriptional activity. Mutants were tested in transfected 293 cells for their ability to activate a co-transfected luciferase reporter gene containing three copies of the Nurr1-binding sites.

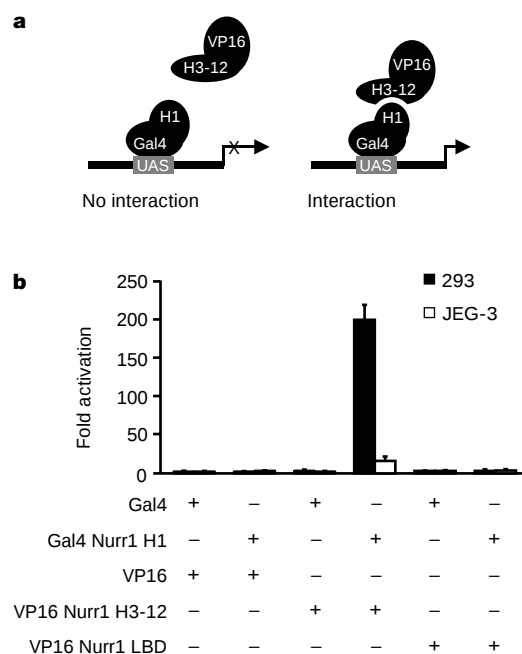


Figure 4 The stability of the Nurr1 LBD correlates with cell-specific transcriptional activity. **a**, Representation of the mammalian two-hybrid LBD assembly assay. H1 of an LBD is fused to the Gal4 yeast transcription factor DBD and the remainder of the LBD (H3–H12) is fused to the strong activation domain (VP16) from herpes simplex virus. **b**, Cell-specific interactions between Nurr1 LBD H1 and the remainder of the Nurr1 LBD (H3–H12) are correlated with LBD-dependent transcriptional activity.

We used mutagenesis to determine whether the coactivator-binding site in Nurr1 could be restored by simply modifying the characteristics of residues participating in the formation of the site. A complete reversal of the reversed charge clamp through the mutations K590E and E422K did not change the activity, and mutations made in this region to make Nurr1 resemble more the coactivator- or corepressor-interacting residues in other NRs did not affect activity (Fig. 3e). The structural observations, together with these mutagenesis data, indicate that there must be an alternative coactivator- or corepressor-binding surface in Nurr1. Thus, H12 most probably has a different role in Nurr1 from its role in other NRs, providing stability to the LBD rather than a direct interaction surface for coregulatory proteins.

Given that the Nurr1 LBD has no ligand-binding cavity and lacks a 'classical' coactivator- or corepressor-binding surface, how is Nurr1 transcriptional activity regulated? As previously observed¹², both Nurr1 and a derivative in which the Nurr1 LBD was fused to the Gal4 yeast transcription factor DNA-binding domain (DBD) were considerably more active in 293 cells than in human chorion carcinoma JEG-3 cells (Supplementary Fig. 1). We used an assembly assay (Fig. 4a) based on the conditional assembly of LBD fragments²⁰ to explore the mechanisms that underlie these differences. For ligand-bound receptors, ligands and corepressors stimulated the interaction of the two LBD fragments and this interaction was detected by the increased transcription of a reporter construct containing Gal4-binding sites. Notably, the difference in transcriptional activity of Nurr1 in the two cell lines strongly correlated with specific assembly of the H1 and H3–H12 fragments of Nurr1 LBD, indicating that this domain is stabilized in 293 cells (Fig. 4b).

As the evidence showed that Nurr1 LBD can adopt different conformations correlating with its transcriptional activity, we investigated whether other signalling pathways might regulate this transition. As both Nurr1 and ligands for the signalling receptor tyrosine kinase Ret are important in midbrain dopamine cells, we

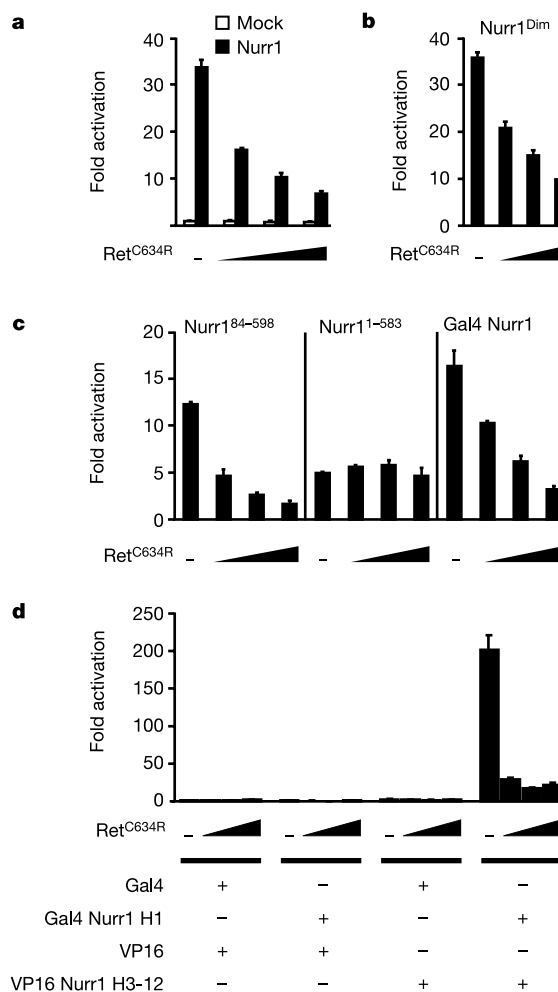


Figure 5 Receptor tyrosine kinase activated intracellular signalling modulates Nurr1 LBD activity and stability. **a**, Constitutively active Ret (Ret^{C634R}) inhibits Nurr1 activation of an NGFI-B response element-containing luciferase reporter in a dose-dependent manner. **b**, **c**, Ret^{C634R}-induced inhibition is equally effective on wild-type Nurr1 and a dimerization-defective Nurr1 mutant (**b**) and affects Nurr1 AF2-dependent transcriptional activity (**c**). **d**, Coexpression of Ret^{C634R} destabilizes Nurr1 intramolecular interactions.

examined the effect of a constitutively active oncogenic derivative of Ret (Ret^{C634R}) in co-transfection experiments. Ret^{C634R} had a profound negative effect on the ability of Nurr1 to activate a reporter gene in 293 cells, but not in JEG-3 cells (Fig. 5a and data not shown). Parallel experiments using a Nurr1 mutant that was unable to form dimers with RXR (Nurr1^{Dim})²¹ showed that this inhibition by Ret^{C634R} was not an indirect effect mediated through RXR (Fig. 5b).

In addition to the AF2 domain, Nurr1 contains an activation function-1 (AF1) domain in the N-terminal region, and a deletion of the N-terminal 83 amino acids of Nurr1 abolishes the activity of AF1 (ref. 12). We found that this deletion mutant was still efficiently repressed by Ret^{C634R}, whereas a mutant containing a deletion of the AF2 domain was not, showing that the AF2 domain, but not the AF1 domain, is inhibited by Ret^{C634R} (Fig. 5c). In addition, when the Nurr1 LBD was tested in a fusion with the Gal4 DBD (Gal4–Nurr1), activation was efficiently repressed by Ret^{C634R}. As Ret signals through, for example, the mitogen-activated protein kinase (MAPK) and phosphatidylinositol-3-OH kinase pathways, we used specific kinase inhibitors to show that inhibition of the Nurr1 AF2 domain is dependent on MAPK activation (Supplementary Fig. 2).

The influence of Ret-dependent signalling on Nurr1 LBD was

assessed using the assembly assay described above. Notably, interaction between the H1 and H3–H12 fragments of Nurr1 LBD was abolished by co-transfection with Ret^{C634R}, showing that the active conformation of Nurr1 LBD is destabilized as a result of Ret signalling (Fig. 5d). Although experiments using the assembly assay provide evidence for structural transitions correlating with transcriptional activity, the mechanism remains to be elucidated. Evidence for a similar regulation of Nurr1 activity and LBD stability was obtained in experiments from neural cell lines (data not shown). Thus, our data show that AF2-dependent transcriptional activity of Nurr1 can be efficiently inhibited by cross-talk through Ret-mediated intracellular signalling and that such a regulatory influence is associated with a structural transition in the Nurr1 LBD.

In conclusion, our data converge to show that Nurr1 activity is correlated with regulated, ligand-independent stabilization of the Nurr1 LBD. The NGFI-B subfamily represents a highly conserved group of NRs²²; thus, the finding that Nurr1 lacks ligand-binding capacity is in accordance with the idea that ligand binding has been independently acquired in several NRs during evolution and that ancestral NRs were not regulated by ligand binding²³. The orphan receptor steroidogenic factor 1 (SF-1) may be an additional example of a NR that is regulated through ligand-independent mechanisms, as previous data have shown that the LBD activity and conformation of SF-1 can be regulated as a result of cross-talk with the MAPK signalling pathway²⁴. In addition, homology modelling and mutagenesis studies of the Rev-Erba subfamily of orphan NRs suggest that they might be examples of ligand-independent transcription factors lacking a cavity for ligand binding, allowing them to constitutively bind corepressor proteins²⁵. Notably, however, our data provide unequivocal evidence of ligand-independent regulation of an orphan NR LBD. □

Methods

Protein preparation

We cloned the human Nurr1 LBD (residues 328–598) by polymerase chain reaction (PCR) into a pET-15b vector (Novagen) with an N-terminal His₆ tag. The protein was expressed in *Escherichia coli* BL21(DE3) cells (Invitrogen), which were grown in either 2YT or minimum medium supplemented with seleno-methionine (SeMet). After two steps of purification by a Ni²⁺ NTA-agarose column (Qiagen) and a Mono-S cation exchange column (Pharmacia), the N-terminal His tag was removed by thrombin. The protein was further purified by another Ni²⁺ NTA-agarose column and a Mono-Q anion exchange column (Pharmacia), and finally using a Superdex 75 gel filtration column (Pharmacia). We concentrated the purified protein to 5–7 mg ml⁻¹ in 20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 2 mM EDTA and 5 mM dithiothreitol for crystallization.

Crystallization

The Nurr1 LBD crystals were grown at 20 °C in either a hanging drop or a sitting drop configuration with 2.5 µl of the protein solution and 2.5 µl of well solution containing 18% (w/v) PEG3350, 0.1 M HEPES (pH 6.5) and 0.2 M KBr. We transferred the crystals into a well solution containing an additional 20% (w/v) of ethylene glycol and then flash-froze them in liquid nitrogen.

Data collection, structure determination and refinement

We collected X-ray diffraction data sets on both a RU-H3RHB generator/RAXIS-IV detector (Rigaku) and on a synchrotron source (beamline 5.0.2) at the Advanced Light Source (ALS) in Berkeley, California. The best diffraction resolution of 2.2 Å was obtained at ALS. All data were integrated using MOSFLM and scaled using SCALA in the CCP4 suite of programs (ref. 26). Crystals grew in two different space groups, P₃₁ and P₃₂₁.

We resolved the Nurr1 LBD structure by two-wavelength (inflection and remote) multiple anomalous dispersion (MAD) phasing with SOLVE²⁷ and SHARP²⁸ using protein incorporating SeMet and the crystal form with space group P₃₂₁. Non-crystallographic symmetry (NCS) averaging using program DM in CCP4 was able to produce a traceable map of high quality. A model was built using the program O²⁹ and refined with REFMAC5 in CCP4. A higher-resolution data set from a crystal belonging to space group P₃₁ was used to complete the final refinement using CNX (ref. 30 and Supplementary Table 1).

The main chain and side chain atoms were visible for residues 363–390 and 399–598 in the final electron density maps for all six molecules in the asymmetric unit in space group P₃₁. Although the first 35 amino acids of the N terminus (residues 328–362) were not visible for five molecules, one molecule did show good quality electron density for residues 334–347. This small fragment served as a centre for the packing of the six molecules.

Chemicals

We purchased PD 98059 and LY 294002 from Calbiochem and dissolved it in dimethyl sulphoxide according to the manufacturer's recommendations.

Cell culture

Human embryonic kidney 293 cells and human chorion carcinoma JEG-3 cells were maintained in Dulbecco's modified Eagle's medium and in minimum essential medium (Invitrogen), respectively. Media were supplemented with 10% heat-inactivated fetal calf serum (FCS) at 37 °C under 5% CO₂ humidified atmosphere.

Transfections and constructs

Transfections were done using Lipofectamine Plus reagent (Invitrogen) according to the manufacturer's recommendations. Typically, 0.5 µg of DNA per well was transfected into cells in 24-well plates for 3–5 h, after which the cells received fresh medium containing FCS (10% final concentration) and chemicals as indicated. Cells were cultured for an additional 20 h and then collected and assayed for luciferase and β-galactosidase activity. Each transfection was done in triplicate or quadruplicate, and each experiment was repeated at least three times. We found that expression of Nurr1 and Nurr1 mutants was equivalent in JEG-3 and 293 cells. Results show the mean ± s.e.m. of representative experiments. The constructs used in this study have been described^{12,21}. Expression vectors pDNA3 Ret^{C634R} and pDNA3 Ret^{Y905F} were a gift from C. Ibanez and contain the coding cDNA sequences of a constitutively active and an inactive mutant of the human c-ret, respectively.

Nurr1 LBD assembly assay

We assessed structural changes in the Nurr1 LBD using a modified version of a described LBD assembly assay²⁰. The expression plasmid pCMX-Gal4–Nurr1–H1 contains the sequence encoding residues 350–398 (corresponding to H1) cloned in frame with the sequence encoding the DBD of the yeast Gal4 transcription factor (residues 1–147). The expression plasmid pCMX-VP16–H3–H12 contains the sequence encoding residues 399–598 (corresponding to H3–H12) cloned in frame with the sequence encoding the activation domain of the VP16 protein from herpes simplex virus.

Received 20 February; accepted 4 April 2003; doi:10.1038/nature01645.

- Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995).
- Giguere, V. Orphan nuclear receptors: from gene to function. *Endocr. Rev.* **20**, 689–725 (1999).
- Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384–391 (1998).
- Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors and lipid physiology: opening the X-files. *Science* **294**, 1866–1870 (2001).
- Kliwer, S. A., Lehmann, J. M. & Willson, T. M. Orphan nuclear receptors: shifting endocrinology into reverse. *Science* **284**, 757–760 (1999).
- Wisely, G. B. *et al.* Hepatocyte nuclear factor 4 is a transcription factor that constitutively binds fatty acids. *Structure* **10**, 1225–1234 (2002).
- Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y. I. & Shoelson, S. E. Crystal structure of the HNF4α ligand binding domain in complex with endogenous fatty acid ligand. *J. Biol. Chem.* **277**, 37973–37976 (2002).
- Law, S. W., Conneely, O. M., DeMayo, F. J. & O'Malley, B. W. Identification of a new brain-specific transcription factor, NURR1. *Mol. Endocrinol.* **6**, 2129–2135 (1992).
- Zetterstrom, R. H. *et al.* Dopamine neuron agenesis in Nurr1-deficient mice. *Science* **276**, 248–250 (1997).
- Saucedo-Cardenas, O. *et al.* Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. *Proc. Natl Acad. Sci. USA* **95**, 4013–4018 (1998).
- Le, W. D. *et al.* Mutations in NR4A2 associated with familial Parkinson disease. *Nature Genet.* **33**, 85–89 (2003).
- Castro, D. S., Arvidsson, M., Bondesson Bolin, M. & Perlmann, T. Activity of the Nurr1 carboxyl-terminal domain depends on cell type and integrity of the activation function 2. *J. Biol. Chem.* **274**, 37483–37490 (1999).
- Wurtz, J. M. *et al.* A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Struct. Biol.* **3**, 87–94 (1996).
- Renaud, J. P. *et al.* Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. *Nature* **378**, 681–689 (1995).
- Brzozowski, A. M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).
- Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-γ. *Nature* **395**, 137–143 (1998).
- Greschik, H. *et al.* Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. *Mol. Cell* **9**, 303–313 (2002).
- Maruyama, K. *et al.* The NGFI-B subfamily of the nuclear receptor superfamily. *Int. J. Oncol.* **12**, 1237–1243 (1998).
- Darimont, B. D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev.* **12**, 3343–3356 (1998).
- Pissios, P., Tzamei, I., Kushner, P. & Moore, D. D. Dynamic stabilization of nuclear receptor ligand binding domains by hormone or corepressor binding. *Mol. Cell* **6**, 245–253 (2000).
- Aarnisalo, P., Kim, C. H., Lee, J. W. & Perlmann, T. Defining requirements for heterodimerization between the retinoid X receptor and the orphan nuclear receptor Nurr1. *J. Biol. Chem.* **277**, 35118–35123 (2002).
- Laudet, V. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**, 207–226 (1997).
- Escriva, H. *et al.* Ligand binding was acquired during evolution of nuclear receptors. *Proc. Natl Acad. Sci. USA* **94**, 6803–6808 (1997).
- Descloux, M., Krylova, I. N., Horn, F., Fletterick, R. J. & Ingraham, H. A. Phosphorylation and intramolecular stabilization of the ligand binding domain in the nuclear receptor steroidogenic factor 1. *Mol. Cell. Biol.* **22**, 7193–7203 (2002).
- Renaud, J. P., Harris, J. M., Downes, M., Burke, L. J. & Muscat, G. E. Structure–function analysis of the Rev-erba and RVR ligand-binding domains reveals a large hydrophobic surface that mediates corepressor binding and a ligand cavity occupied by side chains. *Mol. Endocrinol.* **14**, 700–717 (2000).

26. Collaborative Computational Project Number 4 The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
27. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
28. De la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
30. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **54**, 905–921 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Cao for MS analysis, J. Lehmann, A. Shiau, B. Shan, C. Ibanez, A. Mata and Ö. Wrange for discussions and advice. This work was supported in part by the Göran Gustafsson Foundation, The European Union Research Training Network and the Swedish Foundation for Strategic Research. The Advanced Light Source at the Lawrence Berkeley National Laboratory is supported by the Director, Office of Science, Office of Basic Sciences, Materials Sciences Division, of the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.P.C.W. (walker@tularik.com) or T.P. (thomas.perlmann@licr.ki.se). Atomic coordinates have been deposited in the Protein Data Bank under accession code 1OVL.

SNPs, siRNA and DNAs

Tools for gene and protein handling.



Mouse work: phenotype screening by Taconic.

Phenotyping panels

Taconic Biotechnology www.taconic.com

See your mice in a new light

A series of phenotyping panels from Taconic Biotechnology can be used to help determine the phenotype of gene expression in traditional and genetically altered laboratory mice. The General Screening Panel consists of necropsy, urinalysis, haematology and serum biochemistry tests to provide a view of the fundamental phenotype of the animal. A 'Histology' option adds tissue preservation and pathologist evaluation of the organs surveyed in the necropsy, while the 'Behaviour' option focuses on motor coordination and basic neurologic behaviour of the animal. Finally, the 'Reproductive' option appraises the breeding performance of the strain.

Recombinant Proteinase K

Roche Applied Science

www.roche-applied-science.com

PCR grade now available

Roche's Recombinant PCR-Grade Proteinase K is tested as nuclease-free, function-tested and virtually free of DNA (<10 pg per mg) and other contaminants found in native preparations. It is active over wide pH (4.0–10.0) and temperature (25–50 °C) ranges. This Proteinase K preparation is a high-activity endopeptidase that completely inactivates nucleases during manual or automated purification of DNA or RNA for use in PCR, RT-PCR, cloning, sequencing and many other applications.

Montage PCR m96

Millipore www.millipore.com

Small-volume PCR purification

Millipore's 96-well Montage PCR_μ96 clean-up plates incorporate the company's size-exclusion technology to process PCR reaction volumes from 1 to 150 μl. The microwell plate design is optimized to purify, concentrate and recover PCR products in as little as

20 μl. The 15-minute vacuum-driven protocol removes primers (>99% primer removal) and dNTPs in one step. The plates yield purified PCR product suitable for sensitive genotyping applications, direct sequencing and microarray spotting. In genotyping assays, the Montage plates are particularly useful for MALDI-TOF and CE-based detection methods because of the small reaction volumes. For microarray spotting applications, the plates can concentrate PCR products up to 7.5-fold with high recoveries. Montage plates are compatible with liquid-handling equipment.

ProteomeLab PF 2D

Beckman Coulter www.beckmancoulter.com

Fractionation in two dimensions

The ProteomeLab PF 2D is an automated, two-dimensional fractionation system for high-resolution analysis of complex protein mixtures. In the first dimension, chromatofocusing separates the proteins according to their pI, providing information that can be related to traditional isoelectric focusing techniques. Fractions from the first dimension are collected and automatically injected into a second dimension, which separates on the basis of hydrophobicity. This 'in-solution' approach delivers liquid fractions for further study. The software displays both one- and two-dimensional maps of the expressed protein profiles in easy-to-read formats.

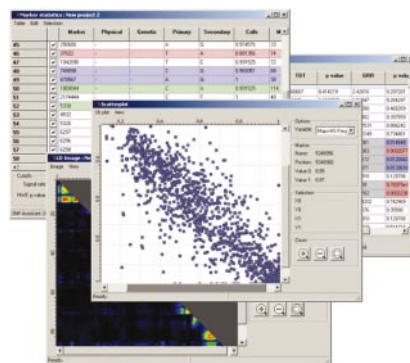
SNP Assistant

Asper Biotech/ BioData

www.asperbio.com/ www.biodata.ee

The generation game

Genotyping specialists Asper Biotech and bioinformatics company BioData are collaborating on a 'next generation' tool for han-



It's a SNP: download a demo from BioData.

ding SNP genotyping data. SNP Assistant is an easy-to-use tool for single nucleotide polymorphism (SNP) management, validation and analysis. Functions include marker and sample validation, case-control and TDT tests, LD calculation and publication level visualization, relationships testing and comparison of datasets. SNP Assistant is not platform-dependent and handles genotyping results from all major technology platforms, enabling groups with different technologies to work in similar environments. A fully functional trial version (time restricted) of the software is available from BioData's website www.biodata.ee.

Platinum PCR Supermix 96

Invitrogen www.invitrogen.com

Try another Taq

Invitrogen's Platinum SuperMix 96 for high-throughput PCR is a ready-to-use, scalable mixture of *Taq* DNA polymerase, anti-*Taq* antibodies, salts, magnesium and dNTPs. The inclusion of monoclonal antibodies that inactivate *Taq* DNA at low temperatures increases specificity. The mixture is pre-aliquoted into 96-well skirted or non-skirted perforated plates.

Amplifluor

Flowgen www.flowgen.co.uk

SNP genotyping with a signal advantage

Flowgen's Amplifluor system for genotyping single nucleotide polymorphisms (SNPs) uses a dual-signal process with two fluorescently labelled primers to distinguish between heterozygous and homozygous alleles. Discrimination of alleles in a single closed tube is achieved with a pair of allele-specific primers with different tail sequences. Depending on the genotype of the sample, a mixed signal of fluorescein and sulphorhodamine fluorescence is observed for a heterozygote, while a single signal from either is observed for a homozygote. Two Amplifluor kits are available — one for assay develop-

ment, which is aimed at new users and has a capacity of 100 reactions, and one for high-throughput screening that can carry out 500 reactions and is designed for investigators who are familiar with the Amplifluor UniPrimer system and SNP genotyping technology. Only one reaction per SNP is needed and the system can use reactions containing as little as 400 pg of genomic DNA.

Dicer siRNA generation kit

Gene Therapy Systems

www.genetherapysystems.com

A little interference

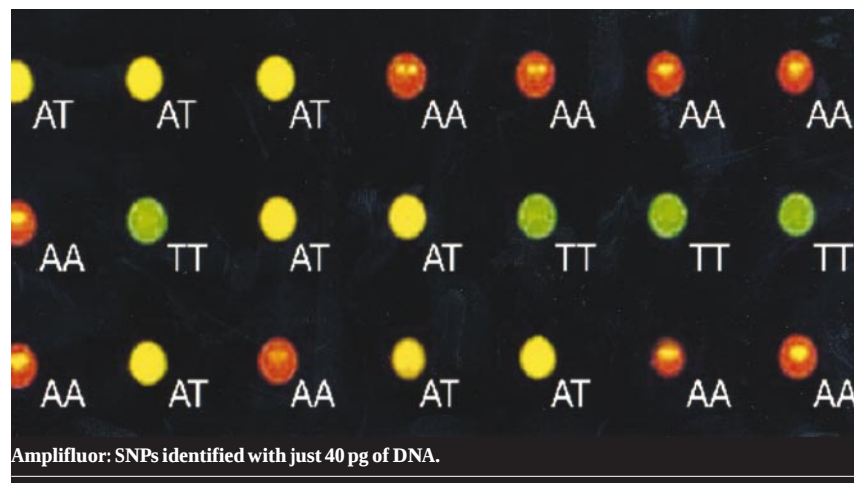
The Dicer siRNA generation kit can be used to generate a large number of small interfering RNAs (siRNAs) from full-length target genes. It eliminates the trial and error associated with synthetic siRNA and siRNA expression vectors by using recombinant human DICER enzyme to cleave *in vitro* transcribed dsRNA templates into a pool of 22 bp siRNAs. The kit contains everything required for preparing double-stranded RNA, RNA cleavage, siRNA purification and transfection.

Perfectprep

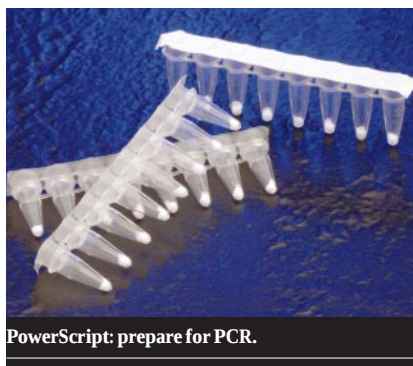
Eppendorf www.eppendorf.com

BACPAC kit

The Perfectprep BAC 96 purification kit from Eppendorf isolates 96 individual BAC, PAC, fosmid or cosmid clones in parallel. Its trapping technology generates high-quality DNA with minimal shearing or chromosomal contamination, optimizing the accuracy of spectrophotometer readings and improving downstream applications such as sequencing (average read length >600 bases), fingerprinting, mutagenesis and PCR. Beginning with the pelleted bacteria, 96 samples can be processed in 60 minutes, or 192 samples in 75 minutes. There is no need for precipitation or resuspension of the purified DNA, so the eluate can be used immediately.



Amplifluor: SNPs identified with just 40 pg of DNA.



PowerScript: prepare for PCR.

BD Sprint PowerScript RT

BD Biosciences Clontech www.clontech.com
Select reverse

This new product combines the BD PowerScript RT reverse transcriptase with a multi-well lyophilized delivery system. BD Sprint PowerScript Single Shots and BD Sprint PowerScript 96 Plate each contain a complete lyophilized master mix of BD PowerScript Reverse Transcriptase, dNTPs and buffer. To use, simply dissolve the lyophilized mixture using PCR-grade water containing diluted primer(s) and RNA and incubate. The makers say these kits provide consistent, RNase-free enzyme preparations ideal for cDNA synthesis and library construction, probe labelling, and all other RT-PCR applications.

Soilmaster

Epicentre www.epicentre.com
Dig for information

The SoilMaster kit can be used to extract PCR-ready DNA from a variety of environmental samples. It uses a hot detergent lysis process combined with a chromatography step that removes organic inhibitors such as humic and fulvic acids that are known to co-extract with DNA from soil and sediment samples. The extracted DNA can be used

with the FailSafe PCR system to amplify bacterial, plant or fungal templates.

Utility cart

Labconco www.labconco.com
Mobile workstation

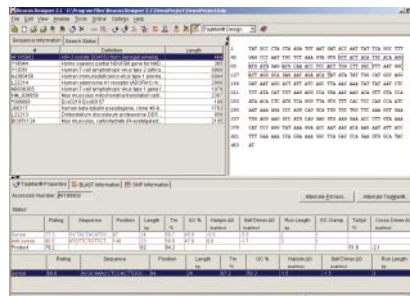
Labconco's utility cart has two open shelves for small electrical equipment and a pair of drawers to hold small labware and supplies. The cart can be used for transporting lab equipment or as a compact work station. It is fitted with a power strip providing 115 volts, 60Hz, 12 amps, with six electrical outlets, a cord bracket and a three-wire cord with plug. The cart supports equipment weighing up to 400 pounds and has lockable casters.

Beacon Designer

Premier Biosoft International
www.premierbiosoft.com

Software for real time PCR

A new version of Beacon Designer has just been released. The program designs primers, TaqMan probes and molecular beacons optimized for quantitative PCR product detection and allele discrimination in multiplex experiments. Beacon Designer uses a sophisticated algorithm to analyse the results of BLAST search and template folding. The results of analysis are used to design highly



Primer suspects: the Beacon Designer screen.

efficient and specific primers. With an intuitive interface and graphical display, the program makes real time assay design and analysis as easy as clicking a button.

KDE GeneSense

Affibody/ InforSense
www.inforsense.com

Functional genomics

Affibody AB and InforSense have jointly introduced KDE GeneSense, a new system for interpreting experimental genomics, transcriptomics and proteomics data. GeneSense couples InforSense's Kensington Discovery Edition (KDE) discovery informatics environment with Affibody's data management, biological annotation and visual representation technology. GeneSense provides researchers with integrated and curated access to relevant information databases; links to KDE's flexible workflow system and powerful analytical tools; interpretation using different knowledge domains/vocabularies (or ontologies); and easy-to-use Virtual Chip and Tree Map visualizers to facilitate the manipulation and display of ontologies.

ProTeam

Tecan www.tecan.com
Single-platform protein processing

ProTeam Advanced Digest is a scalable, flexible one-unit system for automated in-gel or in-solution protein digestion, peptide extraction and MALDI plate-spotting. The system can deliver samples to a range of target plates and can be integrated into Tecan's ProTeam product suite. It is based on Genesis liquid-handling and incorporates the company's DirectSpot technology with TecPrep 96 solid-phase extraction plates to extend sensitivity. Sample volumes as low as 50 nl can be used.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejpn.com

naturejobs

Thinking ahead

Deb Koen tends to have a positive outlook. Vice-president for Career Development Services in Rochester, New York, and a columnist for *The Wall Street Journal's* online careers section, Koen admits that the scientific job market is tough at the moment. But this is still an opportunity, she told attendees at a career fair sponsored by *Naturejobs* and the New York Biotechnology Association this month.

Rather than using the economic conditions as an excuse for not getting a new job, Koen told the meeting that scientists should use them as motivation to hone their skills for basic job hunting and career building. These techniques are often neglected in prosperous times, when scientists' core skills are in demand. But when the job market contracts, they can prove to be invaluable.

The key, Koen said, is being proactive about your career — think of it as a constant quest to find the best match between your skills, values and interests, and those of an employer. Landing a job where these attributes don't match might help to pay the bills, but it will ultimately result in dissatisfaction. To avoid this, Koen suggested that scientists should keep a career journal. Start with your long-term goals, and then identify short-term goals, such as key projects you want to take on or new skills you want to develop. Also, note any trends that are highlighted by the media.

Being proactive goes beyond writing things down, Koen said, it entails following up on your goals. And it means networking — not by introducing yourself to strangers who you hope might help your career, but by being actively involved in your community. It also means seeking feedback — not just before tenure or during performance reviews, and not just from supervisors, but from colleagues and collaborators at different stages of a project. Perhaps then you will be able to make the most of a depressed job market.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Proteomics: the land
of opportunity p566

Using proteins to beat
the recession p568

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Short supply

Improving proteomic techniques will tackle questions in cell biology, signal transduction and clinical research. But workers with the key knowledge in protein biochemistry, mass spectrometry and bioinformatics are hard to find, says Kendall Powell.

Overwhelmed: like many in proteomics, John Bergeron is finding it hard to recruit the skills he needs to deal with the outpouring of protein data.

John Bergeron is being swamped. His team feeds peptide after peptide, isolated from a range of cellular components, into the lab's mass spectrometer, which in turn churns out reams of data. The researchers then scramble to use this information to identify new proteins and the cellular processes in which they are involved. Even though only a quarter or so of the proteins are identified, Bergeron says that his group at McGill University in Montreal, Canada, can barely keep up. Bergeron's experience is not unique — he, along with colleagues in North America and Europe, is finding it difficult to find qualified scientists to help tame the outpouring of protein data.

Proteomics — loosely defined as the study of all of the proteins expressed in an organ, tissue or cell at a given time — presents a challenge vastly more complex than cataloguing genomes. But two sets of papers published last year hint at the scientific rewards and professional opportunities that the field can provide. One showed intricate protein–protein interaction maps from yeast (A.-C. Gavin *et al. Nature* **415**, 141–147 and Y. Ho *et al. Nature* **415**, 180–183; 2002); the other revealed a promising diagnostic marker for ovarian cancer (E. F. Petricoin *et al. Lancet* **359**, 572–577; 2002).

These efforts symbolize two areas ripe for growth in proteomics — the study of protein interactions and pathways, and the search for clinical biomarkers. To take on such projects, research centres need scientists with a thriving proteomics plan for key biological questions and an understanding of how technologies might be used to answer them. In particular, there is a pressing need for mass spectrometrists who can dream up new tools or strategies but who are also classically trained protein biochemists, so that they understand the underlying chemistry of protein separation, identification and characterization.

So where will such scientists come from? Joël Vandekerckhove at Ghent University in Belgium says that across Europe the number of chemistry graduates, a subset of whom will have both the protein-chemistry education and the mass-spectrometry experience

needed for proteomics, is shrinking. Instead of hundreds of graduates, many chemistry departments now produce just a few dozen each year. And many of these tend to head for careers in the pharmaceutical industry. In the United States, 15% fewer biochemistry doctorates were awarded in 2001 compared with 1993, according to a survey by the National Science Foundation.

ACADEMIC SHORTFALL

As a result, recruiting postdocs into academia can be tough, says Simon Gaskell, director of mass spectrometry at the University of Manchester Institute of Science and Technology, UK. "Anyone who can spell mass spectrometry and proteomics can get a job in industry — and probably at 50% higher pay," he says. An increasingly popular way to get a PhD in the field is to work in industry while doing a thesis part-time, he adds.

"Mass spectrometry can be a pretty lucrative environment," says David Muddiman, co-director of the Mayo Proteomics Research Center in Rochester, Minnesota. Students can make up to US\$100,000 a year if they take an analytical chemistry job in industry after graduating, he explains. There are not nearly enough mass-spectrometry groups producing students who really understand the technology, he adds.

Although Muddiman agrees that new technology can take on loads never achieved before — such as his new mass spectrometer, which has a resolution 60 times greater than more conventional models — he cautions against falling in love with a particular technique. "Almost anyone can learn to run a machine like it's a stove and cook a bowl of soup," he says. "But those people are going to be in trouble in 5–10 years." Only mass spectrometrists with a fundamental understanding of the field can move proteomics along by developing the next generation of machines.

Protein bioinformaticians, whose skills include programming and database management, may be the hardest to recruit into academia, as these skills are

Technology is key to proteomics, but researchers should be wary of becoming over-enamoured with equipment, says David Muddiman.

APPLIED BIOSYSTEMS

Reverse migration

Traditionally, scientists move from academia into industry. But in proteomics, there are signs that the traffic is changing direction.

Take Martin Latterich, the former director of proteomics at biotech firm Illumina in San Diego, who will become an associate professor at McGill University in Montreal, Canada, later this year.

Latterich is returning to the challenges of basic proteomics research — an area in which

Illumina is cutting back.

His move signals a change in academic institutions, which are now embracing proteomics technologies that have been proved effective by industry.

Latterich is not unique. Ruedi Aebersold, a proteomics pioneer at the Institute for Systems Biology in Seattle, Washington, says that he has received applications from more researchers in industry than he can possibly employ.

K.P.

much in demand in industry — although there are signs that skilled people from industry are flowing back into academia if their company has foundered (see ‘Reverse migration’, above). As a result, both academic centres and funding agencies allow for higher salary ranges.

While proteomic experts jockey for limited human resources, they all agree that there is room for almost any technical or biological background in the burgeoning field — especially if you can bring a little of both. Because proteins drive all of life’s machinery and represent the vast majority of therapeutic targets, proteomics will be answering questions in basic and clinical research far into the future.

“Even if someone doesn’t want to specialize in proteomics, it’s good to pick up the skills and way of thinking,” says Matthias Mann, a functional-genomics researcher at the University of Southern Denmark in Odense. Mann’s lab illustrates how many skills are needed for a good proteomics lab. Chemists, engineers, biologists and computational experts are organized into either instrumentation or experimentation groups, and these maintain constant contact with each other (see ‘The protein hit squad’, right).

COORDINATED APPROACH

The willingness to work in a team is a prerequisite for scientists who want to work on some of the largest public research efforts in proteomics — perhaps the area in which the most exciting professional and scientific opportunities lie. To try to coordinate these efforts internationally, the Human Proteome Organisation has outlined five initial goals for researchers, including defining the human serum and liver proteomes and standardizing known protein antibodies. The National Heart, Lung, and Blood Institute launched ten proteomics centres around the United States in 2002 with a total funding of \$157 million over seven years.

European resources are also flowing into centres such as the Swedish Human Proteome Resource, a four-year research programme worth 60 million

Swedish kronor (US\$7.6 million) per year. And Britain’s Leukaemia Research Fund set up a proteomics facility at UMIST to give leukaemia researchers access to proteomics experts such as Gaskell who can help them to design cutting-edge projects.

But today’s hot topic is tomorrow’s status quo. Even though last year’s papers on yeast protein–protein interactions and diagnostic markers for ovarian cancer showed promise, the field will have to keep proving it can meet the substantial challenges that lie ahead: perfecting protein arrays to match the quality of DNA arrays, designing better information platforms, and pushing the strategies for identifying complex protein mixtures even further. Because of these new challenges and the ongoing large projects, there should be ample opportunities in proteomics for the foreseeable future — but the best jobs may go to people with a combination of an understanding of biochemistry as well as a background in engineering or computation.

Kendall Powell is a science writer based in Broomfield, Colorado.

Human Proteome Organisation

♦ www.hupo.org

Mayo Proteomics Research Center

♦ www.mayo.edu/research/mprc

Leukaemia Research Fund proteomics facility

♦ www.lrf.umist.ac.uk

Human Proteome Resource

♦ www.biotech.kth.se/molbio/hpr

Genome Canada

♦ www.genomecanada.ca



All together now: Matthias Mann sees integrated teamwork as essential in the proteomics laboratory.

The protein hit squad

People who want to work in Simon Hubbard’s protein bioinformatics group at the University of Manchester Institute of Science and Technology, UK, simply can’t fit the “geeky” stereotype of the computational wizard, he says. This is because his team spends its time designing software to help researchers query public database information in new ways or predict genes from organisms without completed genomes. So everyone must be willing to communicate with a

range of scientists. Apart from programmers, they must also be databank curators, liaison officers to the experimental proteome scientists, and software test pilots.

In fact, Hubbard says that the best part of his job is sitting down with experimental scientists and helping them to identify their proteins. Starting out with a mass of raw data that could represent hundreds of different peptides adds a level of challenge to the job not seen with genomic sequence data, he adds.

K.P.

Growing pains

Current economic conditions are putting a strain on the nascent world of proteomics. But many companies are managing to flourish by carving out their own market niche. Kendall Powell investigates.



Lawrence Cohen sees the in-depth study of proteins as unavoidable.

The economy has been tough on proteomics companies. In the wake of the publication of the draft sequence of the human genome, anything to do with the logical next step — studying protein expression in cells and tissues — looked hot. Companies sprang up to provide technology and protein data to paying customers. And existing firms launched proteomics divisions in the hope that protein data would usher in a new era of drug discovery.

Those expectations have now come down to Earth with a bump, with many small firms either closing shop or scaling back, and several large companies paring down their proteomics divisions. But proteomics is not out of the picture. Companies of all sizes that provide goods and services making direct contributions to drug discovery seem poised to thrive, offering jobs to protein scientists who also have skills in computer science or engineering.

There is a group of companies developing protein-survey technology that is shifting from development to applications and broadening out, says Lawrence Cohen, president and chief executive of Zyomyx, a biotechnology company based in Hayward, California. His firm produced one of the first commercially available protein-array chips, which allows users to test biological samples for a range of proteins from the human immune system. “Proteomics has weathered the storm of the economy because the direct study of proteins is inevitable and imperative to research,” he says.

It is also key to drug discovery, where it will aid the search for valid drug targets and biomarkers for the early detection of disease, as well as help to chart disease-related pathways and protein interactions. As a result, many proteomics companies are starting to explore drug-discovery avenues, with smaller firms gearing themselves up to provide niche services to the larger pharmaceutical companies. In general, what the big drug firms want are proteomics data that fit their suite of drug targets or disease areas, as these could reduce the time spent hunting for active — and safe —



A strong market niche can help proteomics firms to flourish, says Clarissa Desjardins.

lead compounds (see ‘Tight focus’, opposite).

Companies that have carved themselves a niche that appeals to pharmaceutical partners have managed to ride out the economic downturn, says Clarissa Desjardins, founder and an executive vice-president of Caprion Pharmaceuticals in Montreal, Canada. But the key now is to continue to develop that niche. Caprion’s technology probes the proteomes of specific cellular compartments or membranes. Although it is now a reasonably mature technology, the development phase is not necessarily over, Desjardins says. “We want creative people from this field to continue to improve on it so that our partners have more options,” she explains. Caprion also has a growing need for people with scientific backgrounds who can cover business development, law and technology surveillance.

VARIETY SHOW

Cellzome, based in Britain and Germany, also places an emphasis on diversity and teamwork. The company was originally spun off from the European Molecular Biology Laboratory in Heidelberg, Germany. “We liked the international atmosphere there, so we transferred it to Cellzome,” says Gitte Neubauer, co-founder of the company.

Drug development at the firm involves trawling through the human proteome for proteins that bind known drugs or drug candidates. During the next two years, Cellzome plans to add a medicinal chemistry and pharmacology department of about 40 people to bolster its efforts. Neubauer says that initially the firm will hire senior people with industry-level experience, followed by a second round of fresh PhDs or postdocs.

The explosion of interest in proteomics has kept companies that specialize in mass spectrometry and protein separation on their toes. As universities and research institutes add proteomics facilities and encourage exploration in the field, lab-supply companies are adjusting their product lines accordingly.

One such company, Invitrogen in Carlsbad, California, is seeking creative people who can devise

Tight focus

When Hanjo Lim finished his postdoc at the Scripps Research Institute in La Jolla, California, two-and-a-half years ago, he found himself in demand. Coming out of one of the country's premier proteomics labs, Lim had four job offers, two at biotechnology companies and two at major pharmaceutical firms.

He chose Aventis Pharmaceuticals, where he joined as a senior scientist in its proteomics group at Bridgewater,

New Jersey. At the time, biotech and especially proteomics companies were "just not stable" compared with drug firms, he says.

The proteomics group, which focuses on the central nervous system and inflammatory diseases, was just getting started and Lim jumped at the chance to contribute to building a team of mass spectrometry specialists.

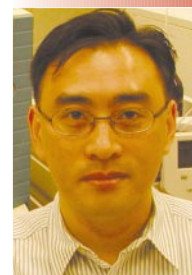
"I was interested in drug discovery rather than the academic field, and if I joined a new

group, I could influence its direction," he says.

Seeing the whole drug-discovery process has taught Lim about bioinformatic and genomic approaches as well. But he notes that there are disadvantages to working in such a product-driven environment. There is no room for technology development, for example, something that Lim hoped he would have time for.

Even so, there is plenty to keep a highly trained proteome scientist

occupied. Lim says that as projects broaden out, his group is moving into mapping protein pathways and finding biomarkers for diagnosing disease. And group members are key advisers to the company on how to outsource proteomic projects. "Proteomics approaches are growing towards more complex systems and are now heavily involved with pathway biology," Lim says. "Coming up with a good experimental plan is not easy." **K.P.**



Hanjo Lim feels that his spell in industry has been rewarding.

new ways of doing something faster, easier or cheaper. "We look for people with solid chemistry or biochemistry training," says Jay Amshey, a vice-president at the company. "If they know why and how something works, then they know enough to make it better." Invitrogen has seen its sales grow over the past two years, and is now hiring PhD-level scientists for all aspects of its business, including manufacturing, sales, business development and law.

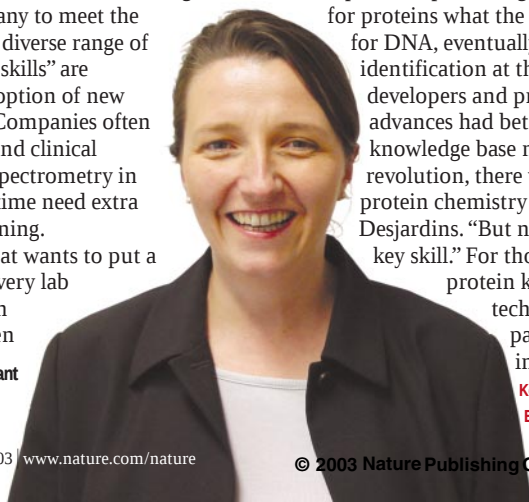
CRITICAL MASS

Applied Biosystems in Foster City, California, is focusing strongly on its mass-spectrometry systems, which are designed with the proteome in mind. "Without question, this is a growth area for us," says David Hicks, the firm's marketing director for proteomics. "But we also want to deliver entirely new workflows for applications and for that we need all the classic talent you would anticipate." That means electrical and mechanical engineers, ion physicists and software developers. Once these systems are developed, the firm needs biochemists and bioinformaticians to test the new products, and scientists to train customers and develop new applications in the field.

Such a broad spread of talent is becoming important for the company to meet the needs of an increasingly diverse range of customers. "Real people skills" are needed to encourage adoption of new technology, Hicks says. Companies often find that cell biologists and clinical researchers using mass spectrometry in proteomics for the first time need extra encouragement and training.

Another company that wants to put a mass spectrometer on every lab bench is also investing in customer care. Ciphergen

Flexibility and variety are important in protein research, according to Gitta Neubauer.



in Fremont, California, has about 50 research scientists around the world who are dedicated to helping clients understand and set up the company's ProteinChip system, which combines protein separation on a chip with time-of-flight mass spectrometry. Ciphergen tends to hire postdocs with training in protein biochemistry for these posts.

Zyomyx also produces a protein chip, which features a number of cell-signalling molecules for testing against a tiny amount of sample fluid. But with the chip and reader system complete, the technology is reasonably mature, says Cohen, so the company is currently seeking people for sales, marketing and technology-support roles.

Cohen believes that the future for protein arrays will see far greater miniaturization and an increased ability to do many experiments in parallel on the same chip. Designing these new arrays, with active proteins tethered to a surface, will require expertise encompassing silicon microfabrication, nanotechnology, microfluidics and detection physics.

The next wave of technology will need to be able to wade through more complex protein mixtures, decipher how many proteins can result from one gene, and tackle 'personalized' molecular medicine. Many experts are predicting that mass spectrometry will do for proteins what the polymerase chain reaction did for DNA, eventually offering rapid protein identification at the bench. But to get there, the developers and promoters of such technical advances had better bulk up their protein knowledge base now. "After the whole genomics revolution, there were lots of gene jockeys and protein chemistry was out of fashion," says Desjardins. "But now it's back and has become a key skill." For those with a combination of

protein knowledge, creativity and the technical skills to make meaningful patterns leap out of the proteome, industry offers a bright future. ■

Kendall Powell is a science writer in Broomfield, Colorado.